Unbalanced Optimal Transport: Geometry and Kantorovich Formulation

Lénaïc Chizat Gabriel Peyré
Bernhard Schmitzer François-Xavier Vialard

Ceremade, Université Paris-Dauphine
{chizat,peyre,schmitzer,vialard}@ceremade.dauphine.fr

November 20, 2015

Abstract

This article presents a new class of "optimal transportation"-like distances between arbitrary positive Radon measures. These distances are defined by two equivalent alternative formulations: (i) a "fluid dynamic" formulation defining the distance as a geodesic distance over the space of measures (ii) a static "Kantorovich" formulation where the distance is the minimum of an optimization program over pairs of couplings describing the transfer (transport, creation and destruction) of mass between two measures. Both formulations are convex optimization problems, and the ability to switch from one to the other depending on the targeted application is a crucial property of our models. Of particular interest is the Wasserstein-Fisher-Rao metric recently introduced independently by [CSPV15, KMV15]. Defined initially through a dynamic formulation, it belongs to this class of metrics and hence automatically benefits from a static Kantorovich formulation. Switching from the initial Eulerian expression of this metric to a Lagrangian point of view provides the generalization of Otto's Riemannian submersion to this new setting, where the group of diffeomorphisms is replaced by a semi-direct product of groups. This Riemannian submersion enables a formal computation of the sectional curvature of the space of densities and the formulation of an equivalent Monge problem.

1 Introduction

Optimal transport is an optimization problem which gives rise to a popular class of metrics between probability distributions. We refer to the monograph of Villani [Vil03] for a detailed overview of optimal transport. A major constraint of the resulting transportation metrics is that they are restricted to measures of equal total mass (e.g. probability distributions). In many applications, there is however a need to compare unnormalized measures, which corresponds to so-called "unbalanced" transportation problems, following the terminology introduced in [Ben03]. Applications of these unbalanced metrics range from

image classification [RGT97, PW08] to the processing of neuronal activation maps [GPC15]. This class of problems requires to precisely quantify the amount of transportation, creation and destruction of mass needed to compare arbitrary positive measures. While several proposals to achieve this goal have been made in the literature (see below for more details), to the best of our knowledge, there lacks a coherent framework that enables to deal with generic measures while preserving both the dynamic and the static perspectives of optimal transport. It is precisely the goal of the present paper to describe such a framework and to explore its main properties.

1.1 Previous Work

In the last few years, there has been an increasing interest in extending optimal transport to the unbalanced setting of measures having non-equal masses.

Dynamic formulations of unbalanced optimal transport. Several models based on the fluid dynamic formulation introduced in [BB00] have been proposed recently [MRSS15, LM13, PR14, PR13]. In these works, a source term is introduced in the continuity equation. They differ in the way this source is penalized or chosen. We refer to [CSPV15] for a detailed overview of these models.

Static formulations of unbalanced optimal transport. Purely static formulations of unbalanced transport are however a longstanding problem. A simple way to address this issue is given in the early work of Kantorovich and Rubinstein [KR58]. The corresponding "Kantorovich norms" were later extended to separable metric spaces by [Han99]. These norms handle mass variations by allowing to drop some mass from each location with a fixed transportation cost. The computation of these norms can in fact be re-casted as an ordinary optimal transport between normalized measures by adding a point "at infinity" where mass can be sent to, as explained by [Gui02]. This reformulation is used in [GPC15] for applications in neuroimaging. A related approach is the so-called optimal partial transport. It was initially proposed in the computer vision literature to perform image retrieval [RGT97, PW08], while its mathematical properties are analyzed in detail by [CM10, Fig10]. As noted in [CSPV15] and recalled in Section 5.1, optimal partial transport is tightly linked to the generalized transport proposed in [PR14, PR13] which allows a dynamic formulation of the optimal partial transport problem. The contributions in [PR14] were inspired by [Ben03] where it is proposed to relax the marginal constraints and to add an L^2 penalization term instead.

Wasserstein-Fisher-Rao metric and relation with recent work. A new metric between measures of non equal masses has recently and independently been proposed by [CSPV15, KMV15]. This new metric interpolates between the Wasserstein W_2 and the Fisher-Rao metrics. It is defined through a dynamic formulation, corresponding formally to a Riemannian metric on the space of

measures, which generalizes the formulation of optimal transport due to Benamou and Brenier [BB00].

In [CSPV15], we proved existence of minimizers in a general setting, presented the limit models (for extreme values of mass creation/destruction cost) and proposed a numerical scheme based on first order proximal splitting methods. We also thoroughly treated the case of two Dirac masses which was the first step towards a Lagrangian description of the model. This metric is the prototypical example for the general framework developed in this article. It thus enjoys both a dynamic formulation and a static one (Sect. 5.2). It is also sufficiently simple so that its geometry (and in particular curvature) can be analyzed in detail, which is useful to get a deep insight about the properties of our generic class of metrics.

1.2 Contribution

The starting point of this article is the Wasserstein-Fisher-Rao (*WF*) metric. For two non-negative densities ρ_0, ρ_1 on a domain $\Omega \subset \mathbb{R}^d$ it is informally obtained by optimizing

$$\mathit{WF}^2(\rho_0, \rho_1) = \inf_{(\rho,v,\alpha)} \int_0^1 \int_\Omega \left(\frac{1}{2}|v(t,x)|^2 + \frac{1}{2}\alpha(t,x)^2 \right) \rho(t,x) \, \mathrm{d}x \, \mathrm{d}t \tag{1.1}$$

where ρ is a time-dependent density, v is a velocity field that describes the movement of mass of ρ and α a scalar field that models local growth and destruction of mass. The triplet (ρ, v, α) must satisfy the following continuity equation with source:

$$\partial_t \rho + \nabla \cdot (\rho\, v) = \rho\, \alpha\,, \qquad \rho(0,\cdot) = \rho_0\,, \qquad \rho(1,\cdot) = \rho_1\,. \tag{1.2}$$

This article presents two sets of contributions. The first one (Section 2) studies in detail the geometry of the *WF* metric and related functionals. It is both of independent interest and serves as a motivation to introduce the second class of contributions. Formula (1.1) is formally interpreted as a particular case of a family of Riemannian metrics on a semi-direct product of groups between diffeomorphisms and scalar functions. The diffeomorphisms account for mass transportation, the scalar fields act as pointwise mass multipliers. The main result of this part is the formal derivation of a submersion of this semi-direct product of groups into the space of positive measures, equipped with a metric from this family (Proposition 2.9). This corresponds to a generalization to *WF* of the Riemannian submersion first introduced by Otto in the optimal transport case [Ott01]. A first application of this result is the computation of the sectional curvature of the *WF* space (Proposition 2.15). A second application is the definition of a Monge-like formulation, i.e. the computation of *WF* in terms of a transport diffeomorphism and a pointwise mass multiplier (Section 2.7).

The second set of contributions introduces and studies a general class of unbalanced optimal transport metrics, which enjoy both a static and a dynamic formulation. We introduce in Section 3 a new Kantorovich-like class of *static*

problems of the form

$$C_K(\rho_0, \rho_1) = \inf_{(\gamma_0,\gamma_1)} \int_{\Omega\times\Omega} c(x, \gamma_0(x,y), y, \gamma_1(x,y))\, \mathrm{d}x\, \mathrm{d}y \tag{1.3}$$

where (γ_0, γ_1) are two 'semi-couplings' between ρ_0 and ρ_1, describing analogously to standard optimal transport, how much mass is transported between any pair x, $y \in \Omega$. Two semi-couplings are required, to be able to describe changes of mass during transport. The function $c(x_0, m_0, x_1, m_1)$ determines the cost of transporting a quantity of mass m_0 from x_0 to a (possibly different) quantity m_1 at x_1. It is a crucial assumption of our approach that $c(x, \cdot, y, \cdot)$ is jointly positively 1-homogeneous and convex in the two mass arguments. This ensures that (1.3) can be rigorously defined as an optimization problem over measures and that the resulting problem is convex. A continuity result and the dual problem are established (Theorems 3.3 and 3.5). Analogous to standard optimal transport, when c induces a metric over pairs of location and mass, then (1.3) defines a metric over non-negative measures (Theorem 3.2).

Then, in Section 4 we look at a family of dynamic problems given by

$$C_C(\rho_0, \rho_1) = \inf_{(\rho,v,\alpha)} \int_0^1 \int_\Omega f\big(x, \rho(t,x), v(t,x)\cdot\rho(t,x), \alpha(t,x)\cdot\rho(t,x)\big)\, \mathrm{d}x\, \mathrm{d}t \tag{1.4}$$

where the infimum is again taken over solutions of (1.2). Here, $f(x, m, v\cdot m, \alpha\cdot m)$ gives the *infinitesimal cost* of moving a chunk of mass m at x in direction v while undergoing an infinitesimal scaling by α. Note that in the two last arguments of f we multiply v and α by m. This corresponds to the velocity $\leftrightarrow$ momentum change of variables proposed in [BB00] to obtain a convex problem. Under suitable assumptions on f (that include (1.1) and go beyond the Riemannian case) we establish equivalence between (1.3) and (1.4) when c is chosen to be the 'pointwise distance' induced by f (Theorem 4.3 and Proposition 4.4).

Finally, we apply those results to two unbalanced optimal transport models. Section 5.1 introduces a dynamic formulation and gives duality results for a family of metrics obtained from the optimal partial transport problem. This is reminiscent of — and generalizes — the results in [PR13, PR14]. The case of the *WF* metric is rigorously discussed in Section 5.2. In Section 5.3 it is shown how standard static optimal transport is obtained as a limit (in the sense of Γ-convergence) of the *WF* metric, thus complementing a previous result of [CSPV15] obtained for dynamic formulations.

While being motivated by Section 2, Sections 3 to 5 are mathematically self-contained. Readers not familiar with infinite dimensional Riemannian geometry do therefore not necessarily need to read Section 2 before proceeding to the rest of the paper.

1.3 Relation with [LMS15a, LMS15b]

After completing this paper, we became aware of the independent work of [LMS15a, LMS15b]. In these two papers the authors develop and study the same class of

"static" transportation-like problems as here. This huge body of work contains many theoretical aspects that we do not cover. For instance measures defined over more general metric spaces are considered, while we work over $\mathbb{R}^d$. The construction of [LMS15a] defines three equivalent static formulations. Their third "homogeneous" formulation is closely related (by a change of variables) to our "semi-couplings" formulation. Their first formulation gives an intuitive and nice interpretation of this class of convex programs as a modification of the original optimal transportation problem where one replaces the hard marginal constraints by soft penalization using Bregman divergences. The dual of this first formulation is related to the dual of our formulation by a logarithmic change of variables. Quite interestingly, the same idea is used in an informal and heuristic way by [FZM+15] for applications in machine learning, where soft marginal constraints is the key to stabilize numerical results. The authors of [LMS15a, LMS15b] study dynamical formulations in the Wasserstein-Fisher-Rao setting (that they call the "Hellinger-Kantorovich" problem). This allows them to make a detailed analysis of the geodesic structure of this space. In contrast, we study a more general class of dynamical problems, but restrict our attention to the equivalence with the static problem. Another original contribution of our work is the proof of the metric structure (in particular the triangular inequality) for static and dynamic formulations when the underlying cost over the cone manifold $\Omega \times \mathbb{R}^+$ is related to a distance. Lastly, our geometric study of Section 2, and in particular the Riemannian submersion structure, the explicit sectional curvature computation and the Monge problem appear to be original contributions. Note that along these lines, the work of [LMS15a] proves a lower bound on the Alexandrov curvature in the *WF* case, which in particular allows these authors to state sufficient condition for the *WF* space to have positive curvature. In a smooth setting, we find similar results in Proposition 2.15 and Corollary 2.16.

1.4 Preliminaries and Notation

We denote by $C(X)$ the Banach space of real valued continuous functions on a compact set $X \subset \mathbb{R}^d$ endowed with the sup norm topology. Its topological dual is identified with the set of Radon measures, denoted by $\mathcal{M}(X)$ and the dual norm on $\mathcal{M}(X)$ is the total variation, denoted by $|\cdot|_{TV}$. Another useful topology on $\mathcal{M}(X)$ is the weak* topology arising from this duality: a sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ weak* converges towards $\mu \in \mathcal{M}(X)$ if and only if for all $u \in C(X)$, $\lim_{n\to+\infty} \int_X u \,\mathrm{d}\mu_n = \int_X u \,\mathrm{d}\mu$. According to that topology, $C(X)$ and $\mathcal{M}(X)$ are topologically paired spaces (the elements of each space can be identified with the continuous linear forms on the other), this is a standard setting in convex analysis. We also use the following notations:

- $\mathcal{M}_+(X)$ is the space of nonnegative Radon measures, $\mathcal{M}_+^{ac}(X)$, the subset of absolutely continuous measures w.r.t. the Lebesgue measure and $\mathcal{M}_+^{at}(X)$ the subset of purely atomic measures.

- For M a given manifold, TM denotes the tangent bundle of M and T_pM the tangent space at a point $p \in M$.

- $\mathrm{Dens}(X)$ the set of finite Radon measures that have smooth positive density w.r.t. a reference volume form ν;
- $\mu \ll \nu$ means that the $\mathbb{R}^m$-valued measure μ is absolutely continuous w.r.t. the positive measure ν. We denote by $\frac{\mu}{\nu} \in (L^1(X,\nu))^m$ the density of μ with respect to ν.
- For a (possibly vector) measure μ, $|\mu| \in \mathcal{M}_+(X)$ is its variation;
- For a map $\varphi : M \mapsto N$ between two manifolds M, N, $T\varphi$ denotes the tangent map of φ. For a given Riemannian metric g on N, the pull-back of g by φ is denoted by $\varphi^* g$ and defined by $(\varphi^* g)(x)(v_x, v_x) \stackrel{\text{def.}}{=} g(\varphi(x))(T_x\varphi(v_x), T_x\varphi(v_x))$.
- $T_\# \mu$ is the image measure of μ through the measurable map $T : X_1 \to X_2$, also called the pushforward measure. It is given by $T_\#\mu(A_2) \stackrel{\text{def.}}{=} \mu(T^{-1}(A_2))$; When $X_1 = X_2$ is a manifold and T is a diffeomorphism, we denote $T_\#$ by T_*.
- δ_x is a Dirac measure of mass 1 located at the point x;
- $\iota_{\mathcal{C}}$ is the (convex) indicator function of a convex set $\mathcal{C}$ which takes the value 0 on $\mathcal{C}$ and $+\infty$ everywhere else;
- If (E, E') are topologically paired spaces and $f : E \to \mathbb{R} \cup \{+\infty\}$ is a convex function, f^* is its Legendre transform i.e. for $y \in E'$, $f^*(y) \stackrel{\text{def.}}{=} \sup_{x\in E} \langle x, y\rangle - f(x)$.
- For $n \in \mathbb{N}$ and a tuple of distinct indices $(i_1, \dots, i_k)$, $i_l \in \{0, \dots, n-1\}$ the map
$$\mathrm{Proj}_{i_1,\dots,i_k} : \Omega^n \to \Omega^k \tag{1.5}$$
denotes the canonical projection from Ω^n onto the factors given by the tuple $(i_1, \dots, i_k)$.
- The truncated cosine is defined by $\overline{\cos} : z \mapsto \cos(|z| \wedge \frac{\pi}{2})$.

2 The Geometric Formulation in a Riemannian setting

This section is focussed on Riemannian generalizations of the Wasserstein-Fisher-Rao (*WF*) metric. The *WF* distance, as informally defined in (1.1) over the space of Radon measures on Ω is the motivating example for the geometric formulation of Section 2 and also a simple example for which an equivalent static formulation exists in the setting of Section 5.2.

This is a prototypical example of metrics over densities that can be written as
$$G^2(\rho_0, \rho_1) = \inf_{\rho,v,\alpha} \int_0^1 \left(\frac{1}{2} \int_\Omega g(x)((v,\alpha),(v,\alpha)) \mathrm{d}\rho_t(x) \right) \mathrm{d}t \tag{2.1}$$

under the same constraints where $g(x)$ is a scalar product on $T_x\Omega \times \mathbb{R}$ where the two factors $v = v(t,x)$ and $\alpha = \alpha(t,x)$ represent the velocity field and the growth rate. Note that *WF* is obtained from (2.1) by choosing $g(x)((v,\alpha),(v,\alpha)) = 2\,f(1,v,\alpha) = |v|^2 + \delta^2\,\alpha^2$. We will see in Section 2.2 that this family of metrics is exactly all the Riemannian metrics that satisfies the homogeneity condition formulated in Definition 2.1. This homogeneity condition appears compulsory in order to properly define the metric on the space of Radon measure via means of convex analysis (see Section 4).

While Section 1.2 defines the *WF* metric over a bounded domain $\Omega \subset \mathbb{R}^d$, we will assume in the rest of the section that Ω is a compact manifold, possibly with smooth boundary.

2.1 Otto's Riemannian Submersion: Eulerian and Lagrangian Formulations

Standard optimal transport consists of moving one distribution of mass to another while minimizing a transportation cost, which is an optimization problem originally formulated in Lagrangian (static) coordinates. In [BB00], the authors introduced a convex Eulerian formulation (dynamic) which enables the natural generalization proposed in [CSPV15, KMV15]. The link between the static and dynamic formulation is made clear using Otto's Riemannian submersion [Ott01] which emphasizes the idea of a group action on the space of probability densities. More precisely, let Ω be a compact manifold and $\mathrm{Diff}(\Omega)$ be the group of smooth diffeomorphisms of Ω and $\mathrm{Dens}_p(\Omega)$ be the set of probability measures that have smooth positive density with respect to a reference volume measure ν. We consider such a probability density denoted by ρ_0. Otto proved that the map

$$
\begin{aligned}
&\Pi : \mathrm{Diff}(\Omega) \to \mathrm{Dens}_p(\Omega) \\
&\Pi(\varphi) = \varphi_* \rho_0
\end{aligned}
$$

is a Riemannian submersion of the metric $L^2(\rho_0)$ on $\mathrm{Diff}(\Omega)$ to the Wasserstein W_2 metric on $\mathrm{Dens}_p(\Omega)$. Therefore, the geodesic problem on $\mathrm{Dens}_p(\Omega)$ can be reformulated on the group $\mathrm{Diff}(\Omega)$ as the Monge problem,

$$
W_2(\rho_0,\rho_1)^2 \stackrel{\text{def.}}{=} \inf_{\varphi \in \mathrm{Diff}(\Omega)} \left\{ \int_\Omega \|\varphi(x) - x\|^2 \rho_0(x)\,\mathrm{d}\nu(x) \,:\, \varphi_*\rho_0 = \rho_1 \right\} . \tag{2.2}
$$

For an overview on the geometric formulation of optimal transport, we refer the reader to [KW08] and to [Del09] for a more detailed presentation.

2.2 Admissible Riemannian Metrics

In our setting, where mass is not only moved but also changed, the group acting on a mass particle has to include a mass rescaling action in addition to the transport action. Let us introduce informally the Lagrangian formulation of the

continuity constraint with source associated to (1.2). Let $m(t)\,\delta_{x(t)}$ describe a particle of mass $m(t)$ at point $x(t)$. The continuity constraint with source reads

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}x(t) = v(t,x(t)) \\ \frac{\mathrm{d}}{\mathrm{d}t}m(t) = \alpha(t,x(t))\,m(t)\,. \end{cases} \tag{2.3}$$

This states that the mass is dragged along by the vector field v and simultaneously undergoes a growth process at rate α. These equations also represent the infinitesimal action of the group which is described in more abstract terms in the next section.

Another important object is the metric used to measure spatial and mass changes. Note that the metric g introduced in (2.1) defines a unique Riemannian metric on the product space $\Omega \times \mathbb{R}_+^*$ which transforms homogeneously under pointwise multiplication. Since the pointwise multiplication will be used, it is natural to consider this product space as a trivial principal fiber bundle where the structure group is $\mathbb{R}_+^*$ under multiplication. In Section 4 we will prove an equivalence result between dynamic and static formulations for a general cost function (see Definition 4.2) which reduces, in the Riemannian case, to this type of metrics. We therefore define the following admissible class of metrics:

Definition 2.1 (Admissible Riemannian metrics)**.** A smooth Riemannian metric g on $\Omega\times\mathbb{R}_+^*$ will be said to be *admissible* if the family of maps $\Psi_\lambda : (\Omega\times\mathbb{R}_+^*, \lambda\, g) \to (\Omega \times \mathbb{R}_+^*, g)$, $\lambda > 0$, defined by $\Psi_\lambda(x,m) = (x, \lambda\, m)$ are isometries.

Under such a metric, the metric completion of $\Omega \times \mathbb{R}_+^*$ is the cone over Ω which we now define.

Definition 2.2 (Cone)**.** The cone over Ω denoted by $\mathrm{Cone}(\Omega)$ is the quotient space $(\Omega \times \mathbb{R}_+) \,/\, (\Omega \times \{0\})$. The apex of the cone $\Omega \times \{0\}$ will be denoted by $\mathcal{S}$.

Proposition 2.1. *The metric completion of* $(\Omega\times\mathbb{R}_+^*, g)$ *for an admissible metric* g *is the cone* $\mathrm{Cone}(\Omega)$.

Proof. By the definition we have that $d((x,m),(x,m/2)) = m^{1/2} d((x,1),(x,1/2))$. Taking $m = 1/2^k$, we get

$$d((x,1),(x,0)) \leq \sum_{k\geq 0} d((x,1/2^k),(x,1/2^{k+1})) = \frac{\sqrt{2}}{\sqrt{2}-1} d((x,1),(x,1/2))\,. \tag{2.4}$$

Moreover, $d((x,m),(y,m)) = md((x,1),(y,1))$ therefore, (x,m) and (y,m) have the same limit when m goes to 0. This limit is the apex of the cone as defined in Definition 2.2. Then, the set $\mathcal{S} \cup (\Omega\times]0,m_0])$ is compact since Ω is assumed so.

Now, consider a Cauchy sequence (x_n, m_n) for the distance induced by the *admissible* metric. It implies that m_n is bounded above since $d((x_n,m_n),(x_n,0)) = m_n d((x_n,1),(x_n,0))$ and thus (x_n,m_n) has an accumulation point in the cone. ☐

Remark. Note that Definition 2.1 and Proposition 2.1 are also valid for Finsler metrics under minor changes.

Proposition 2.2. *Any admissible metric on $\Omega \times \mathbb{R}_+^*$ is completely defined by its restriction to $\Omega \times \{1\}$. There exist $\tilde{g}$ a metric on Ω, $a \in T^*\Omega$ a 1-form and b a positive function on Ω such that*

$$g(x,m) = m\,\tilde{g}(x) + a(x)\,\mathrm{d}m + b(x)\frac{\mathrm{d}m^2}{m}\,. \tag{2.5}$$

Proof. For an *admissible* Riemannian metric one has

$$\lambda\, g(x,m)((v_x,v_m),(v_x,v_m)) = g(x,\lambda\, m)((v_x,\lambda\, v_m),(v_x,\lambda\, v_m)) \tag{2.6}$$

for all $\lambda > 0$, $(x,m) \in \Omega \times \mathbb{R}_+^*$ and $(v_x,v_m) \in T_{(x,m)}\Omega \times \mathbb{R}_+^*$. As a consequence, we have that

$$g(x,m)((v_x,v_m),(v_x,v_m)) = m\, g(x,1)((v_x,v_m/m),(v_x,v_m/m))\,. \tag{2.7}$$

Expanding the terms,

$$\begin{aligned} g(x,m)((v_x,v_m),(v_x,v_m)) = mg(x,1)((v_x,0),(v_x,0)) + 2g(x,1)((v_x,0),(0,v_m) \\ + \frac{1}{m}g(x,1)((0,v_m),(0,v_m)\,, \end{aligned} \tag{2.8}$$

we obtain the desired decomposition and the fact that the *admissible* metric is completely defined by $g(x,1)$. ☐

Note that for a given 1-form $a \in T^*\Omega$ and b a positive function on Ω, the formula (2.5) defines a metric if and only if its determinant is everywhere positive.

We will also use the short notation

$$g(x)(v_x,\alpha) \stackrel{\text{def.}}{=} g(x,1)((v_x,v_m/m),(v_x,v_m/m)) \tag{2.9}$$

with $\alpha = v_m/m$ as it was introduced in (2.1).

Using a square root change of variables, this type of metrics can be related to (generalized) Riemannian cones. Recall that a Riemannian cone (see [Gal79, BBI01] for instance) on a Riemannian manifold (Ω,h) is the manifold $\Omega \times \mathbb{R}_+^*$ endowed with the cone metric $g_c \stackrel{\text{def.}}{=} m^2 h + \mathrm{d}m^2$. The change of variables $\Psi : (x,m) \mapsto (x,\sqrt{m})$ gives $\Psi^* g_c = mh + \frac{1}{4m}\mathrm{d}m^2$, which is the *admissible* metric associated with the initial Wasserstein-Fisher-Rao metric (1.1). This type of metrics is well-known and we summarize hereafter some important properties:

Proposition 2.3. *Let (Ω,g) be a complete Riemannian manifold and consider $\Omega \times \mathbb{R}_+^*$ with the admissible metric defined by $m\,g + \frac{1}{4m}\mathrm{d}m^2$ for $(x,m) \in \Omega \times \mathbb{R}_+^*$. For a given vector field X on Ω, define its lift on $\Omega \times \mathbb{R}_+^*$ by $\tilde{X} = (X,0)$ and denote by e the vector field defined by $\frac{\partial}{\partial m}$. This Riemannian manifold has the following properties:*

1. *Its curvature tensor satisfies* $R(\tilde{X}, e) = 0$ *and*

$$R(\tilde{X}, \tilde{Y})\tilde{Z} = (R_g(X, Y)Z - g(Y, Z)X + g(X, Z)Y, 0) \tag{2.10}$$

where R_g *denotes the curvature tensor of* (Ω, g).

2. *The distance on* $\mathrm{Cone}(\Omega)$ *is*

$$d\left((x_0, m_0), (x_1, m_1)\right) = \left[m_0 + m_1 - 2\sqrt{m_0 m_1} \cos\left(d(x_0, x_1) \wedge \pi\right)\right]^{1/2}. \tag{2.11}$$

Proof. The proof of the first point is in [Gal79] and the second point can be found in [BBI01]. Note that the square root change of variables $\Psi : (x, m) \mapsto (x, \sqrt{m})$ is needed for the application of these results. □

Note that (as remarked in [Gal79]), for any geodesic c on Ω parametrized with unit speed, the map $\phi : \mathbb{C} \setminus \mathbb{R}^- \to \Omega \times \mathbb{R}^*_+$ defined by $\phi(me^{i\theta}) = (c(\theta), m^2)$ is a local isometry.

Corollary 2.4. *If* (Ω, g) *has sectional curvature greater than* 1, *then* $(\Omega \times \mathbb{R}^*_+, m\, g + \frac{1}{4m}\mathrm{d}m^2)$ *has non-negative sectional curvature and more precisely for* X, Y *two orthornormal vector fields on* Ω,

$$K(\tilde{X}, \tilde{Y}) = \frac{1}{m^2}(K_g(X, Y) - 1) \tag{2.12}$$

where K *and* K_g *denote respectively the sectional curvatures of* $\Omega \times \mathbb{R}^*_+$ *and* Ω.

Although the Riemannian cone over a segment in $\mathbb{R}$ is locally flat, the curvature still concentrates at the apex of the cone.

In view of applications, it is of practical interest to classify, at least locally, the space of *admissible* metrics. The first important remark is that the metric associated with the *WF* model on $\mathbb{R}$ of the form $mg + \frac{1}{m}\mathrm{d}m^2$ is flat. It is a particular case of *admissible* metrics that are diagonal, which we define hereafter.

Definition 2.3. A diagonal *admissible* metric is a metric on $\Omega \times \mathbb{R}^*_+$ that can be written as $m\, g + \frac{c}{m}\mathrm{d}m^2$ where g and c are respectively a metric and a positive function on Ω.

It is possible to exhibit *admissible* diagonal metrics that have non zero sectional curvature when $\Omega \subset \mathbb{R}$. Therefore, *admissible* diagonal metrics are not isometric to the standard Riemannian cone. The next proposition gives a characterization of *admissible* metrics that can be diagonalized by a fiber bundle isomorphism [Mic08, Section 17] and it shows that there is a correspondence between diagonal metrics and exact 1-forms on Ω given by principal fiber bundle isomorphisms (see [Mic08, Section 18.6] for instance).

Proposition 2.5. *Any admissible metric* $h(x, m) = mh(x) + a(x)\mathrm{d}m + b(x)\frac{\mathrm{d}m^2}{m}$ *on* $\Omega \times \mathbb{R}^*_+$ *is the pull back of a diagonal admissible metric by a principal fiber bundle isomorphism if and only if* $\frac{a(x)}{b(x)}$ *is an exact* 1*-form. More precisely, there exist positive functions* c, λ *on* Ω *and a metric* g *on* Ω *such that* $\Phi^*(mg + \frac{c}{m}\mathrm{d}m^2) = h$ *where* $\Phi(x, t) = (x, \lambda(x)t)$.

See Appendix B for a proof. In particular, the proof shows that in the space of *admissible* metrics, locally diagonalizable metrics are in correspondence with closed 1-forms. Thus, the space of *admissible* metrics is strictly bigger than the space of standard cone metrics and even strictly bigger than diagonal metrics. This statements are to be understood "up to fiber bundle isomorphisms" which respect the decomposition between space and mass.

2.3 A Semi-direct Product of Groups

As mentioned before, we denote by ν a volume form on Ω. We first denote $\Lambda(\Omega) \stackrel{\text{def.}}{=} \{\lambda \in C^\infty(\Omega, \mathbb{R}) : \lambda > 0\}$ which is a group under the pointwise multiplication and recall that $\mathrm{Dens}(\Omega)$ is the set of finite Radon measures that have smooth positive density w.r.t. the reference measure ν. We first define a group morphism from $\mathrm{Diff}(\Omega)$ into the automorphism group of $\Lambda(\Omega)$, $\Psi : \mathrm{Diff}(\Omega) \to \mathrm{Aut}(\Lambda(\Omega))$ by $\Psi(\varphi) : \lambda \mapsto \varphi^{-1} \cdot \lambda$ where $\varphi \cdot \lambda \stackrel{\text{def.}}{=} \lambda \circ \varphi^{-1}$ is the usual left action of the group of diffeomorphisms on the space of functions. The map Ψ is an antihomomorphism since it reverses the order of the action. The associated semi-direct product is well-defined and it will be denoted by $\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$. We recall the following properties for $\varphi_1, \varphi_2 \in \mathrm{Diff}(\Omega)$ and $\lambda_1, \lambda_2 \in \Lambda(\Omega)$,

$$(\varphi_1, \lambda_1) \cdot (\varphi_2, \lambda_2) = (\varphi_1 \circ \varphi_2, (\varphi_2^{-1} \cdot \lambda_1)\lambda_2) \tag{2.13}$$

$$(\varphi_1, \lambda_1)^{-1} = (\varphi_1^{-1}, \varphi_1 \cdot \lambda_1^{-1})\,. \tag{2.14}$$

Note that this is not the usual definition of a semi-direct product of groups but it is isomorphic to it. We chose this definition in order to get the following left-action:

Proposition 2.6 (Left action)**.** *The map π defined by*

$$\begin{aligned} &\pi : (\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)) \times \mathrm{Dens}(\Omega) \mapsto \mathrm{Dens}(\Omega) \\ &\pi\left((\varphi, \lambda), \rho\right) \stackrel{\text{def.}}{=} (\varphi \cdot \lambda)\, \varphi_* \rho = \varphi_*(\lambda \rho) \end{aligned}$$

is a left-action of the group $\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$ on the space of densities.

Proof. This can be checked by the following elementary calculation:

$$\begin{aligned} \pi\left((\varphi_1, \lambda_1) \cdot (\varphi_2, \lambda_2), \rho\right) &= \pi\left((\varphi_1 \circ \varphi_2, (\varphi_2^{-1} \cdot \lambda_1)\lambda_2), \rho\right) \\ &= (\varphi_1 \circ \varphi_2) \cdot ((\varphi_2^{-1} \cdot \lambda_1)\lambda_2)(\varphi_1 \circ \varphi_2) \cdot \rho \\ &= (\varphi_1 \cdot \lambda_1)(\varphi_2 \cdot \lambda_2)(\varphi_1 \circ \varphi_2) \cdot \rho \\ &= \pi\left((\varphi_1, \lambda_1), \pi\left((\varphi_2, \lambda_2), \rho\right)\right)\,. \end{aligned}$$

The identity element in $\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$ is $(\mathrm{Id}, 1)$ and one trivially has:

$$\pi\left((\mathrm{Id}, 1), \rho\right) = (\mathrm{Id} \cdot 1)\, \mathrm{Id}_* \rho = \rho\,.$$

□

2.4 Generalization of Otto's Riemannian Submersion

In this section, we define useful notions to obtain the generalization of Otto's result. The next definition is a simple change of variables on the tangent space of the group. It represents the change between Lagrangian and Eulerian point of view.

Definition 2.4 (Right-trivialization)**.** Let H be a group and a smooth manifold at the same time, possibly of infinite dimensions, the right-reduction of TH is the bundle isomorphism $\tau : TH \mapsto H \times T_{\mathrm{Id}}H$ defined by $\tau(h, X_h) \overset{\text{def.}}{=} (h, T\mathcal{R}_{h^{-1}}X_h)$, where X_h is a tangent vector at point h and $\mathcal{R}_{h^{-1}} : H \to H$ is the right multiplication by h^{-1}, namely, $\mathcal{R}_{h^{-1}}(f) = fh^{-1}$ for all $f \in H$.

In the finite dimensional case, we would have chosen to work with H a Lie group, however, in infinite dimensions, being a Lie group is too restrictive as shown by Omori [Omo78]. For instance, in fluid dynamics, the right-trivialized tangent vector $X_h \cdot h^{-1}$ is the spatial or Eulerian velocity (the vector field) and X_h is the Lagrangian velocity. Note that most of the time, this right-trivialization map is continuous but not differentiable due to a loss of smoothness of the right composition (see [EM70]).

Example 2.1. For the semi-direct product of groups defined above, we have

$$\tau((\varphi,\lambda),(X_\varphi,X_\lambda)) = ((\varphi,\lambda),(X_\varphi \circ \varphi^{-1}, \varphi \cdot (X_\lambda \lambda^{-1}))) \,, \tag{2.15}$$

or equivalently,

$$\tau((\varphi,\lambda),(X_\varphi,X_\lambda)) = ((\varphi,\lambda),(X_\varphi \circ \varphi^{-1}, (X_\lambda \lambda^{-1}) \circ \varphi^{-1})) \,. \tag{2.16}$$

We will denote by (v,α) an element of the tangent space of $T_{(\mathrm{Id},1)} \operatorname{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$. Any path on the group can be parametrized by its initial point and its right-trivialized tangent vector. The reconstruction equation reads

$$\begin{cases} \partial_t \varphi(t,x) = v(t,\varphi(t,x)) \\ \partial_t \lambda(t,x) = \alpha(t,\varphi(t,x))\lambda(t,x) \end{cases} \tag{2.17}$$

for given initial conditions $\varphi(0,x)$ and $\lambda(0,x)$. Note that this system recovers equation (2.3).

We state without proof a result that will be needed in the Kantorovich formulation and which is a straightforward consequence of a Cauchy-Lipschitz result, whose proof can be found in [O'R97].

Proposition 2.7. *If $v \in L^1([0,T], W^{1,\infty}(\Omega))$, then the first equation in (2.17) has a unique solution in $W^{1,\infty}(\Omega)$. If, in addition, $\alpha \in L^\infty(\Omega)$, then the system (2.17) has a unique solution.*

We also need the notion of infinitesimal action associated with a group action.

Definition 2.5 (Infinitesimal action)**.** For a smooth left action of H on a manifold M, the infinitesimal action is the map $T_{\mathrm{Id}}H \times M \mapsto TM$defined by

$$\xi \cdot q \stackrel{\text{def.}}{=} \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \exp(\xi t) \cdot q \in T_q M \tag{2.18}$$

where $\exp(\xi t)$ is the solution to $\dot{h} = \xi \cdot h$ and $h(0) = \mathrm{Id}$.

Example 2.2. For $\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$ the application of the definition gives $(v, \alpha) \cdot \rho = -\nabla \cdot (v\rho) + \alpha\rho$. Indeed, one has

$$(\varphi(t), \lambda(t)) \cdot \rho = \mathrm{Jac}(\varphi(t)^{-1})(\lambda(t)\rho) \circ \varphi^{-1}(t)\,.$$

First recall that $\partial_t \varphi(t) = v \circ \varphi(t)$ and $\partial_t \lambda = \alpha\lambda(t)$. Once evaluated at time $t = 0$ where $\varphi(0) = \mathrm{Id}$ and $\lambda(0) = 1$, the differentiation with respect to φ gives $-\nabla \cdot (v\rho)$ and the second term $\alpha\rho$ is given by the differentiation with respect to λ.

We now recall a standard construction to obtain Riemannian submersions from a transitive group action in the situation where the isotropy subgroups are conjugate to each others. The next proposition is a reformulation of [Mic08, Claim of Section 29.21] which is concerned with the finite dimensional case. We formally apply the result in our context which is infinite dimensional.

Proposition 2.8. *Suppose that a smooth left action of Lie group H on a manifold M is transitive and such that for every $\rho \in M$, the infinitesimal action $\xi \mapsto \xi \cdot \rho$ is a surjective map. Let $\rho_0 \in M$ and a Riemannian metric G on H that can be written as:*

$$G(h)(X_h, X_h) = g(h \cdot \rho_0)(X_h \cdot h^{-1}, X_h \cdot h^{-1}) \tag{2.19}$$

for $g(h \cdot \rho_0)$ an inner product on $T_{\mathrm{Id}}H$. Let $X_\rho \in T_\rho M$ be a tangent vector at point $h \cdot \rho_0 = \rho \in M$, we define the Riemannian metric $\overline{g}$ on M by

$$\overline{g}(\rho)(X_\rho, X_\rho) \stackrel{\text{def.}}{=} \min_{\xi \in T_{\mathrm{Id}}H} g(\rho)(\xi, \xi) \text{ under the constraint } X_\rho = \xi \cdot \rho\,. \tag{2.20}$$

where $\xi = X_h \cdot h^{-1}$

Then, the map $\pi_0 : H \mapsto M$ defined by $\pi_0(h) = h \cdot \rho_0$ is a Riemannian submersion of the metric G on H to the metric $\overline{g}$ on M.

Note that, by hypothesis, the infinitesimal action is supposed to be surjective, therefore the optimization set is not empty and it needs to be checked that the infimum is attained (in infinite dimensions). This will be done in Proposition 2.12.

Note also that the submersion can be rewritten as the quotient map from H into the space of right-cosets $\pi : H \to H_0 \backslash H$ where H_0 is the isotropy subgroup of ρ_0 in H. Therefore, other fibers of the submersion are right-cosets of the subgroup H_0 in H.

We now apply this construction to the action of the semi-direct product of group onto the space of densities in order to retrieve the class of *WF* metrics:

We choose a reference smooth density ρ_0 and for a general *admissible* metric we define the Riemannian metric we will use on $\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$ by, denoting $\varphi \cdot \lambda\, \varphi_* \rho_0$ by ρ and using the same notation for the infinitesimal action in (2.2),

$$
\begin{aligned}
G(\varphi,\lambda)\left((X_\varphi, X_\lambda),(X_\varphi, X_\lambda)\right) &= \frac{1}{2}\int_\Omega g(x)((v(x),\alpha(x)),(v(x),\alpha(x)))\rho(x)\,\mathrm{d}x \\
&= \frac{1}{2}\int_\Omega g((X_\varphi \circ \varphi^{-1},(X_\lambda \lambda^{-1})\circ\varphi^{-1}),(X_\varphi\circ\varphi^{-1},(X_\lambda\lambda^{-1})\circ\varphi^{-1}))\rho\,\mathrm{d}x
\end{aligned} \tag{2.21}
$$

where $(X_\varphi, X_\lambda) \in T_{(\varphi,\lambda)}\,\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$ is a tangent vector at (φ,λ). Recall that $g(x)$ is an inner product on $T_x\Omega \times \mathbb{R}$ that depends smoothly on x as defined in (2.9). The initial *WF* model reads

$$
\begin{aligned}
G(\varphi,\lambda)\left((X_\varphi, X_\lambda),(X_\varphi, X_\lambda)\right) &= \frac{1}{2}\int_\Omega |v(x)|^2\rho(x)\mathrm{d}x + \frac{\delta^2}{2}\int_\Omega \alpha(x)^2\rho(x)\mathrm{d}x \\
&= \frac{1}{2}\int_\Omega |X_\varphi\circ\varphi^{-1}|^2 \varphi\cdot\lambda\,\varphi_*\rho_0(x)\mathrm{d}x + \frac{\delta^2}{2}\int_\Omega \left(\varphi\cdot(X_\lambda\lambda^{-1})\right)^2\varphi\cdot\lambda\,\varphi_*\rho_0(x)\mathrm{d}x\,.
\end{aligned} \tag{2.22}
$$

At a formal level, we thus get, for ρ_0 a measure of finite mass and which has a smooth density w.r.t. the reference measure ν and for the general metric G defined above:

Proposition 2.9 (Riemannian Submersion)**.** *Let* $\rho_0 \in \mathrm{Dens}(\Omega)$ *and* $\pi_0 : \mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega) \mapsto \mathrm{Dens}(\Omega)$ *be the map defined by* $\pi_0(\varphi,\lambda) \overset{\text{def.}}{=} \varphi_*(\lambda\rho_0)$*.*

Then, the map π_0 *is formally a Riemannian submersion of the metric* G *on the group* $\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$ *to the metric WF on the space of densities* $\mathrm{Dens}(\Omega)$*.*

This proposition is formal in the sense that we do not know if the metrics G and *WF* and the map π_0 are smooth or not for some well chosen topologies and if the horizontal lift is well defined. We address the smoothness of G in the next section.

2.5 Curvature of $\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$

In this section, we are interested in curvature properties of the space $\mathrm{Diff}(\Omega) \ltimes_\Psi \Lambda(\Omega)$. Since we want to work in a smooth setting, we will work in a stronger topology on the group than the one defined by the metric G. Therefore, we will use the definition of a weak Riemannian metric [EM70, Section 9] that we recall below.

Definition 2.6 (Weak metric)**.** Let X be a Hilbert manifold modeled on a Hilbert space H. A weak Riemannian metric g on X is a smooth map $x \in X \mapsto g(x)$ into the space of positive definite bilinear forms on T_xX.

Note that the inner product on T_x need not define the topology on T_xX since it can be weaker than the scalar product on H. In order to give a rigorous meaning to the next lemma, we will work on the group of Sobolev diffeomorphisms $\mathrm{Diff}^s(\Omega)$

for $s > d/2+1$ and $\Lambda^s(\Omega) \stackrel{\text{def.}}{=} \{f \in H^s(\Omega) \,:\, f > 0\}$. We refer to [BV13] for a more detailed presentation of $\mathrm{Diff}^s(\Omega)$ and we only recall that it is contained in the group of C^1 diffeomorphisms of Ω. We first prove a lemma that shows that the metric G is a weak Riemannian metric on $\mathrm{Diff}^s(\Omega) \ltimes_\Psi \Lambda^s(\Omega)$.

Lemma 2.10. *On $\mathrm{Diff}^s(\Omega) \ltimes_\Psi \Lambda^s(\Omega)$, one has*

$$G(\varphi,\lambda)\,((X_\varphi, X_\lambda),(X_\varphi, X_\lambda)) = \frac{1}{2}\int_\Omega g(\varphi(x),\lambda(x))((X_\varphi(x), X_\lambda(x)),(X_\varphi(x), X_\lambda(x)))\rho_0(x)\,\mathrm{d}\nu(x)\,, \quad (2.23)$$

which is a weak Riemannian metric.

Remark. Since G is only a weak Riemannian metric, the Levi-Civita connection does not necessarily exists as explained in [EM70] or in [MMM13].

Proof. In the definition of the metric G, we make the change of variables by φ^{-1}, which is allowed since $\varphi \in \mathrm{Diff}^s(\Omega)$ and the definition of an *admissible* metric to obtain formula (2.23). Since Ω is compact, λ attains its strictly positive lower bound. In addition, using the fact that g is a smooth function and $H^s(\Omega)$ is a Hilbert algebra, the metric is also smooth. □

Note that the formulation (2.23) shows that this metric is an L^2 metric on the space of functions from Ω into $\Omega \times \mathbb{R}_+^*$ endowed with the Riemanian metric g. Then, the group $\mathrm{Diff}^s(\Omega) \ltimes_\Psi \Lambda^s(\Omega)$ is an open subset of $H^s(\Omega, \Omega \times \mathbb{R}_+^*)$. These functional spaces have been studied in [EM70] as manifolds of mappings and they prove, in particular, the existence of a Levi-Civita connection for $\mathrm{Diff}^s(\Omega)$ endowed with an L^2 metric.

Theorem 2.11 (Sectional curvature of the group)**.** *Let $\Omega \times \mathbb{R}_+^*$ endowed with an admissible Riemannian metric g and ρ be a density on Ω. Let X, Y be two smooth vector fields on $\mathrm{Diff}^s(\Omega) \ltimes_\Psi \Lambda^s(\Omega)$ which are orthogonal for the $L^2(\Omega,\rho)$ scalar product on $H^s(\Omega, \Omega\times\mathbb{R}_+^*)$. Denoting $\mathcal{K}_p$ the curvature tensor of G at point $p = (\varphi,\lambda)$, one has*

$$\mathcal{K}_p(X_p, Y_p) = \int_\Omega K_{p(x)}(X_p(x), Y_p(x))(|X_p(x)|^2|| Y_p(x)|^2 - \langle X_p(x), Y_p(x)\rangle)\,\rho(x)\mathrm{d}\nu(x) \quad (2.24)$$

where $X_p \stackrel{\text{def.}}{=} X(p)$ and $\langle\cdot,\cdot\rangle$ and $|\cdot|$ stands for the metric g. In addition, K_y denotes the sectional curvature of $(\Omega \times \mathbb{R}_+^*, g)$ at the point $y \in \Omega \times \mathbb{R}_+^*$.

Proof. Since Ω is compact, the appendix in [FG89] can be applied and it gives the result. □

Remark. It can be useful for the understanding of Formula (2.24) to recall some facts that can be found in [EM70] or [Mis93] and [FG89]. Let us denote $M \stackrel{\text{def.}}{=} \Omega$ and $N \stackrel{\text{def.}}{=} \Omega \times \mathbb{R}_+^*$. The first step of the proof of 2.11 is the existence of the Levi-Civita connection which is a direct adaptation of [EM70, Section 9]. Denoting

$\pi_1 : TTN \to TN$ be the canonical projection. As recalled in [EM70, Section 2], one has, for M, N smooth compact manifold with smooth boundaries,

$$T_f H^s(M,N) = \{g \in H^s(M,TN) \,:\, \pi_1 \circ g = f\}\,,$$

and

$$TTH^s(M,N) = \{Y \in H^s(M,TTN) \,:\, \pi_1 \circ Y \in TH^s(M,TN)\}\,.$$

Denote by $\mathcal{K}$ the connector associated with the Levi-Civita connection of g, a careful detailed presentation of connectors can be found in [Bes78]. The Levi-Civita connection $\tilde{\nabla}$ on $H^s(M,N)$ endowed with the L^2 metric with respect to the metric g on N and the volume form μ_0 on M is given by:

$$\tilde{\nabla}_X Y = \mathcal{K} \circ TY \circ X\,. \tag{2.25}$$

The result on the curvature tensor can be deduced from these facts.

2.6 Curvature of Dens(Ω)

This section is concerned with the formal computation of the curvature of Dens(Ω). The *WF* metric can be proven to be a weak Riemannian metric on the space of densities of H^s regularity. However, the Levi-Civita does not exist. These two facts are proven in Appendix A. Moreover, in this context, the submersion defined in Section 2.4 is not smooth due to a loss of regularity. The rest of the section will thus consist in formal computations.

In order to apply O'Neill's formula, we need to compute the horizontal lift of a vector field on Dens(Ω). In this case of a left action, there is a natural extension of the horizontal lift of a tangent vector at point $\rho \in$ Dens(Ω). Recall that the horizontal lift is defined by formula (2.20). The following proposition is straightforward:

Proposition 2.12 (Horizontal lift)**.** *Let $\rho \in$ Dens(Ω) be a smooth density and $X_\rho \in C^\infty(\Omega, \mathbb{R})$ be a smooth function that represents a tangent vector at the density ρ. The horizontal lift at $(\mathrm{Id}, 1)$ of X_ρ is given by $(\nabla\Phi, \Phi)$ where Φ is the solution to the elliptic partial differential equation:*

$$-\nabla \cdot (\rho \nabla \Phi) + \Phi \rho = X_\rho\,, \tag{2.26}$$

with homogeneous Neumann boundary conditions.

Proof. Using the formula (2.20), the horizontal lift of the tangent vector X_ρ is given by the minimization of the norm of a tangent vector (v, α) at $(\mathrm{Id}, 1)$

$$\inf_{v,\alpha} \frac{1}{2} \int_\Omega g((v,\alpha),(v,\alpha))\rho \,\mathrm{d}\nu(x)\,, \tag{2.27}$$

under the constraint $-\nabla\cdot(\rho v) + \alpha\rho = X_\rho$. This is a standard projection problem for the space $L^2(\Omega, \mathbb{R}^d) \times L^2(\Omega, \mathbb{R})$ endowed with the scalar product defined in (2.27) (recall that ρ is positive on a compact manifold). The existence of a

minimizer is thus guaranteed and there exists a Lagrange multiplier $\Phi \in L^2(\Omega, \mathbb{R})$ such that the minimizer will be of the form $(\nabla\Phi, \Phi)$. Therefore, the solution to the elliptic partial differential equation (2.26) is the solution. By elliptic regularity theory, the solution Φ is smooth. $\square$

In order to compute the curvature, we only need to evaluate it on any horizontal lift that projects to X_ρ at point ρ. There is a natural lift in this situation given by the right-invariant vector field on $\mathrm{Diff}^s(\Omega) \ltimes_\Psi \Lambda^s(\Omega)$.

Definition 2.7. Let $(v, \alpha) \in T_{(\mathrm{Id},1)} \mathrm{Diff}^s(\Omega) \ltimes_\Psi \Lambda^s(\Omega)$ be a tangent vector. The associated right-invariant vector field $\xi_{(v,\alpha)}$ is given by

$$\xi_{(v,\alpha)}(\varphi, \lambda) \stackrel{\text{def.}}{=} ((\varphi, \lambda), (v \circ \varphi, \alpha \circ \varphi \, \lambda)) \,. \tag{2.28}$$

Remark. Note that, due to the loss of smoothness of the right composition, this vector field $\xi_{(v,\alpha)}$ is smooth for the H^s topology if and only if (v, α) is C^∞.

Last, we need the Lie bracket of the horizontal vector fields on the group. In the case of right-invariant vector fields on the group, their Lie bracket is the right-invariant vector field associated with the Lie bracket on the manifold. Therefore, we have:

Proposition 2.13. *Let (v_1, α_1) and (v_2, α_2) be two tangent vectors at identity. Then,*

$$[(v_1, \alpha_1), (v_2, \alpha_2)] = ([v_1, v_2], \nabla\alpha_1 \cdot v_2 - \nabla\alpha_2 \cdot v_1) \,, \tag{2.29}$$

where $[v_1, v_2]$ denotes the Lie bracket on vector fields, and therefore,

$$[\xi_{(v_1,\alpha_1)}, \xi_{(v_2,\alpha_2)}] = \xi_{([v_1,v_2], \nabla\alpha_1 \cdot v_2 - \nabla\alpha_2 \cdot v_1)} \,. \tag{2.30}$$

Thus, applying this formula to horizontal vector fields gives

Corollary 2.14. *Let ρ be a smooth density and X_1, X_2 be two tangent vectors at ρ, Φ_1, Φ_2 be the corresponding solutions of (2.26). We then have*

$$[\xi_{(\nabla\Phi_1,\Phi_1)}, \xi_{(\nabla\Phi_2,\Phi_2)}] = \xi_{([\nabla\Phi_1,\nabla\Phi_2],0)} \,. \tag{2.31}$$

We formally apply O'Neill's formula to obtain a similar result to standard optimal transport. This formal computation could be probably made more rigorous following [Lot08] in a smooth context or following [AG13]. It is however not possible to apply the O'Neill formula developed in [MMM13] in an infinite dimensional setting due to the lack of regularity of the submersion and the non-existence of the Levi-Civita connection as shown in Appendix A.2.

Proposition 2.15 (Sectional curvature of *WF*). *Let ρ be a smooth density and X_1, X_2 be two orthonormal tangent vectors at ρ and*

$$Z_1 = (\nabla\Phi_1, \Phi_1) = \xi_1((\mathrm{Id}, 1)), Z_2 = (\nabla\Phi_2, \Phi_2) = \xi_2((\mathrm{Id}, 1))$$

their corresponding right-invariant horizontal lifts on the group. If O'Neill's formula can be applied, the sectional curvature of $\mathrm{Dens}(\Omega)$ *at point* ρ *is given by*

$$K(\rho)(X_1, X_2) = \int_\Omega k(x,1)(Z_1(x), Z_2(x)) w(Z_1(x), Z_2(x)) \rho(x) \mathrm{d}\nu(x) \\ + \frac{3}{4} \left\| [Z_1, Z_2]^V \right\|^2 \quad (2.32)$$

where

$$w(Z_1(x), Z_2(x)) = g(x)(Z_1(x), Z_1(x)) g(x)(Z_2(x), Z_2(x)) - g(x)(Z_1(x), Z_2(x))^2$$

and $[Z_1, Z_2]^V$ *denotes the vertical projection of* $[Z_1, Z_2]$ *at identity and* $\|\cdot\|$ *denotes the norm at identity.*

Proof. This is the application of O'Neill's formula [Lan99, Corollary 6.2] and Proposition 2.9 at the reference density ρ together with Theorem 2.11 on the vector fields ξ_1, ξ_2 on the group. □

Remark. It is important to insist on the fact that we only compute the "local" sectional curvature. We have seen that the geometry of space and mass is that of a Riemannian cone, in which the curvature concentrates at the apex, although the Riemannian cone can be locally flat. Obviously, in this infinite dimensional context, the sectional curvature only gives information in smooth neighborhoods of the density.

Corollary 2.16. *Let* (Ω, g) *be a compact Riemannian manifold of sectional curvature bounded below by* 1*, then the sectional curvature of* $(\mathrm{Dens}(\Omega), \mathit{WF})$ *is non-negative.*

We stated this Corollary due to the interest in displacement convexity of the Boltzmann entropy (see [Vil09, Corollary 17.19]). In the case where Ω is the (Euclidean) sphere, the Riemannian cone with the standard metric $\mathrm{Cone}(\Omega)$ is flat and a finer characterisation of null sectional curvature for *WF* can be given. Namely, the sectional curvature vanishes if and only if $\nabla\phi_1$ and $\nabla\phi_2$ are commuting vector fields on the sphere.

2.7 The Corresponding Monge Formulation

In this section, we discuss a formal Monge formulation that will motivate the development of the corresponding Kantorovich formulation in Section 3.

Let us recall an important property of a Riemannian submersion

$$\pi : (M, g_M) \mapsto (B, g_B)\,.$$

Every horizontal lift of a geodesic on the base space B is a geodesic in M. In turn, given any two points $(p, q) \in B$, any length minimizing geodesic between the fibers $\pi^{-1}(p)$ and $\pi^{-1}(q)$ projects down onto a length minimizing geodesic on B between p and q. From the point of view of applications, it can be either

interesting to compute the geodesic downstairs and then lift it up horizontally or going the other way.

In the context of this generalized optimal transport model, the Riemannian submersion property is shown in Proposition 2.9. Moreover, the metric on $\mathrm{Diff}(\Omega) \ltimes_{\Psi} \Lambda(\Omega)$ is an L^2 metric as proven in the lemma 2.10, which is particularly simple. Therefore, it is possible to formulate the corresponding Monge problem

$$WF(\rho_0, \rho_1) = \inf_{(\varphi,\lambda)} \left\{ \|(\varphi, \lambda) - (\mathrm{Id}, 1)\|_{L^2(\rho_0)} \,:\, \varphi_*(\lambda \rho_0) = \rho_1 \right\} . \tag{2.33}$$

Let us denote by d the distance on $\Omega \times \mathbb{R}_+^*$ associated with an *admissible* Riemannian metric, then we have:

$$\|(\varphi, \lambda) - (\mathrm{Id}, 1)\|^2_{L^2(\rho_0)} = \int_\Omega d\left((\varphi(x), \lambda), (x, 1)\right)^2 \, \rho_0(x) \, \mathrm{d}\nu(x) . \tag{2.34}$$

In the case of a standard Riemannian cone with the metric $m\, g + \frac{1}{4m} \mathrm{d}m^2$, Proposition 2.3 gives the explicit expression of the distance which gives

$$\|(\varphi, \lambda) - (\mathrm{Id}, 1)\|^2_{L^2(\rho_0)} = \int_\Omega 1 + \lambda - 2\sqrt{\lambda} \cos\left(d(\varphi(x), x) \wedge \pi\right) \, \rho_0(x) \, \mathrm{d}\nu(x) . \tag{2.35}$$

From a variational calculus point of view, it is customary to pass from the Monge formulation to its relaxation. So, instead of making rigorous statements on this Monge formulation, we will directly work on the Kantorovich formulation in the next sections. However, in the next sections, we do not restrict our study to Riemannian costs and we extend it to general dynamical costs that are introduced in Definition 4.2. As a motivation for this generalization we can mention the case of L^p norms for $p \geq 1$, namely:

$$G(\varphi, \lambda)\, (X_\varphi, X_\lambda) \;=\; \frac{1}{p} \int_\Omega |v(x)|^p \rho(x) \mathrm{d}\nu(x) \;+\; \frac{1}{p} \int_\Omega |\alpha(x)|^p \rho(x) \mathrm{d}\nu(x) . \tag{2.36}$$

In Lagrangian coordinates, this gives rise to an L^p metric on $\Omega \times \mathbb{R}_+^*$, namely if $y = m^{1/p}$, then $g(x, y)(v_x, v_y) = (y\|v_x\|)^p + |v_y|^p$. Note that the distance induced by the L^p norm does not correspond to the standard L^p norm on the Euclidean cone. However, it is possible to retrieve the standard L^p norm on the cone in the general setting of Section 4, by pulling it back in spherical coordinates.

3 Static Kantorovich Formulations

Inspired by the Monge formulation of Section 2.7 which emphasizes the importance of the distance on the space $\Omega \times \mathbb{R}_+^*$, we now propose a general Kantorovich problem in Definition 3.3 and its dual formulation in Theorem 3.5. It includes as a particular case the relaxation of the Monge formulation. This Kantorovich formulation does not need the cost on the product space to be related to a distance, yet, when it is the case and under mild assumptions described below, the Kantorovich formulation defines a distance on $\mathcal{M}_+(X)$ as shown in Theorem 3.2.

3.1 Definitions

In what follows, Ω is a compact set in $\mathbb{R}^d$, x typically refers to a point in Ω and m to a mass. We first require some properties on the cost function.

Definition 3.1 (Cost function)**.** In the sequel, a *cost function* is a function

$$c: \begin{array}{ccc} (\Omega \times [0, +\infty[)^2 & \to & [0, +\infty] \\ (x_0, m_0), (x_1, m_1) & \mapsto & c(x_0, m_0, x_1, m_1) \end{array}$$

which is l.s.c. in all its arguments and jointly sublinear in (m_0, m_1).

A sublinear function is by definition a positively 1-homogeneous and subadditive function, or equivalently, a positively 1-homogeneous and convex function. The joint sub-additivity of c in (m_0, m_1) guarantees that it is always better to send mass from one point to another in one single chunk. In order to allow for variations of mass, we need to adapt the constraint set of standard optimal transport by introducing two "semi-couplings", which are relaxed couplings with only one marginal being fixed.

Definition 3.2 (Semi-couplings)**.** For two marginals $\rho_0, \rho_1 \in \mathcal{M}_+(\Omega)$, the set of semi-couplings is

$$\Gamma(\rho_0, \rho_1) \stackrel{\text{def.}}{=} \left\{(\gamma_0, \gamma_1) \in \left(\mathcal{M}_+(\Omega^2)\right)^2 : (\mathrm{Proj}_0)_\# \gamma_0 = \rho_0,\ (\mathrm{Proj}_1)_\# \gamma_1 = \rho_1\right\}. \tag{3.1}$$

Informally, $\gamma_0(x, y)$ represents the amount of mass that is taken from ρ_0 at point x and is then transported to an (possibly different, to account for creation/destruction) amount of mass $\gamma_1(x, y)$ at point y of ρ_1. These semi-couplings allow us to formulate a novel static Kantorovich formulation of unbalanced optimal transport as follows.

Definition 3.3 (Unbalanced Kantorovich Problem)**.** For a cost function c we introduce the functional

$$J_K(\gamma_0, \gamma_1) \stackrel{\text{def.}}{=} \int_{\Omega^2} c\left(x, \tfrac{\gamma_0}{\gamma}, y, \tfrac{\gamma_1}{\gamma}\right) \mathrm{d}\gamma(x, y), \tag{3.2}$$

where $\gamma \in \mathcal{M}_+(\Omega^2)$ is any measure such that $\gamma_0, \gamma_1 \ll \gamma$. This functional is well-defined since c is jointly 1-homogeneous w.r.t. the mass variables (see Definition 3.1). The corresponding optimization problem is

$$C_K(\rho_0, \rho_1) \stackrel{\text{def.}}{=} \inf_{(\gamma_0, \gamma_1) \in \Gamma(\rho_0, \rho_1)} J_K(\gamma_0, \gamma_1). \tag{3.3}$$

Proposition 3.1. *If c is a* cost function *then a minimizer for $C_K(\rho_0, \rho_1)$ exists.*

Proof. Let $\hat{\Omega}$ be an open set such that $\Omega \subset \hat{\Omega}$ and extend c as $+\infty$ if either x or y belongs to $\hat{\Omega} \setminus \Omega$. Then by virtue of [AFP00, Theorem 2.38], J_K is weakly* l.s.c. on $\mathcal{M}(\hat{\Omega}^2)$ and *a fortiori* on $\mathcal{M}(\Omega^2)$. Since Ω is compact and the marginals ρ_0, ρ_1 have finite mass, $\Gamma(\rho_0, \rho_1)$ is tight and thus weakly* pre-compact. It is also closed so $\Gamma(\rho_0, \rho_1)$ is weakly* compact and any minimizing sequence admits a cluster point which is a minimizer (the minimum is not assumed to be finite). $\square$

Example 3.1. Standard optimal transport problems with a nonnegative, l.s.c. cost $\tilde{c}$ are retrieved as particular cases, by taking

$$c(x_0, m_0, x_1, m_1) = \begin{cases} m_0 \cdot \tilde{c}(x, y) & \text{if } m_0 = m_1\,, \\ +\infty & \text{otherwise.} \end{cases}$$

Other examples, in particular the *WF*-metric, are studied in more detail in Section 5.

3.2 Properties

A central property of optimal transport is that it can be used to lift a metric from the base space Ω to the space of probability measures over Ω (cf. [Vil09, Chapter 6]). We now show that this extends to our framework.

Theorem 3.2 (Metric)**.** *Let c be a cost function such that, for some $p \in [1, +\infty[$*

$$(x_0, m_0), (x_1, m_1) \mapsto c(x_0, m_0, x_1, m_1)^{1/p} \tag{3.4}$$

is a (pseudo-)metric on $\mathrm{Cone}(\Omega)$*. Then $C_K(\cdot,\cdot)^{1/p}$ defines a (pseudo-)metric on $\mathcal{M}_+(\Omega)$.*

Remark. The space $\mathrm{Cone}(\Omega)$ is the space $\Omega \times [0, +\infty[$, where all points with zero mass $\Omega \times \{0\}$ are identified as defined in Definition 2.2. Note also that the metric $C_K(\cdot,\cdot)^{1/p}$ may take the value $+\infty$ if c does so and the denomination "pseudo-metric" is sometimes preferred for such objects.

Proof. First, symmetry and non-negativity are inherited from c. Moreover,

$$\begin{aligned} &[C_K(\gamma_0, \gamma_1) = 0] \Leftrightarrow \\ &\qquad [\exists (\gamma_0, \gamma_1) \in \Gamma(\rho_0, \rho_1) : (\gamma_0 = \gamma_1) \text{ and } (x = y\ \gamma_0\text{-a.e.})] \\ &\qquad\qquad \Leftrightarrow [\rho_0 = \rho_1]. \end{aligned}$$

It remains to show the triangle inequality. Fix ρ_0, ρ_1, $\rho_2 \in \mathcal{M}_+(\Omega)$. Take two pairs of minimizers for (3.3)

$$(\gamma_0^{01}, \gamma_1^{01}) \in \Gamma(\rho_0, \rho_1)\,, \qquad\qquad (\gamma_0^{12}, \gamma_1^{12}) \in \Gamma(\rho_1, \rho_2)\,,$$

and let $\mu \in \mathcal{M}_+(\Omega)$ be such that $(\mathrm{Proj}_1)_\#(\gamma_0^{01}, \gamma_1^{01}) \ll \mu$ and $(\mathrm{Proj}_0)_\#(\gamma_0^{12}, \gamma_1^{12}) \ll \mu$. Denote by $(\gamma_i^{01}|\mu)(x|y)$ the disintegration of γ_i^{01} along the second factor w.r.t. μ. More precisely, for all $y \in \Omega$, $(\gamma_i^{01}|\mu)(\cdot|y) \in \mathcal{M}_+(\Omega)$ and it holds, for all f measurable on Ω^2,

$$\int_{\Omega^2} f \,\mathrm{d}\gamma_i^{01} = \int_\Omega \left(\int_\Omega f(x, y)\,\mathrm{d}(\gamma_i^{01}|\mu)(x|y) \right) \mathrm{d}\mu(y)$$

and analogously for $(\gamma_i^{12}|\mu)(z|y)$ along the first factor for $i = 0, 1$. Write $\frac{\rho_1}{\mu}(y)$ for the density of ρ_1 w.r.t. μ. We combine the optimal semi-couplings in a suitable

way to define γ_0, γ_1, $\hat{\gamma} \in \mathcal{M}(\Omega^3)$ (via disintegration w.r.t. μ along the second factor):

$$(\gamma_0|\mu)(x,z|y) \stackrel{\text{def.}}{=} \begin{cases} \frac{(\gamma_0^{01}|\mu)(x|y)\otimes(\gamma_0^{12}|\mu)(z|y)}{\frac{\rho_1}{\mu}(y)} & \text{if } \frac{\rho_1}{\mu}(y)>0, \\ (\gamma_0^{01}|\mu)(x|y)\otimes\delta_y(z) & \text{otherwise,} \end{cases}$$
$$(\gamma_1|\mu)(x,z|y) \stackrel{\text{def.}}{=} \begin{cases} \frac{(\gamma_1^{01}|\mu)(x|y)\otimes(\gamma_1^{12}|\mu)(z|y)}{\frac{\rho_1}{\mu}(y)} & \text{if } \frac{\rho_1}{\mu}(y)>0, \\ \delta_y(x)\otimes(\gamma_1^{12}|\mu)(z|y) & \text{otherwise,} \end{cases}$$
$$(\hat{\gamma}|\mu)(x,z|y) \stackrel{\text{def.}}{=} \begin{cases} \frac{(\gamma_1^{01}|\mu)(x|y)\otimes(\gamma_0^{12}|\mu)(z|y)}{\frac{\rho_1}{\mu}(y)} & \text{if } \frac{\rho_1}{\mu}(y)>0, \\ 0 & \text{otherwise.} \end{cases}$$

The interpretation of γ_0 is that all mass that leaves x towards y, according to $\gamma_0^{01}(x,y)$, is distributed over the third factor according to $\gamma_0^{12}(y,z)$. In case the mass disappears at y, it is simply "dropped" as δ_y on the third factor. Then γ_1 is built analogously for the incoming masses and $\hat{\gamma}$ is a combination of incoming and outgoing masses. For $i=0,1$ let $\gamma_i^{02} \stackrel{\text{def.}}{=} (\mathrm{Proj}_{02})_\#\gamma_i$ and note that, by construction, $(\gamma_0^{02},\gamma_1^{02}) \in \Gamma(\rho_0,\rho_2)$. In the rest of the proof, for an improved readability, when writing the functional the "dummy" measure γ such that $\gamma_0,\gamma_1 \ll \gamma$ is considered as implicit and we write

$$\int_{\Omega^2} c\big(x,\gamma_0(x,y),y,\gamma_1(x,y)\big)\,\mathrm{d}x\,\mathrm{d}y \quad \text{for} \quad \int_{\Omega^2} c\left(x,\tfrac{\gamma_0}{\gamma},y,\tfrac{\gamma_1}{\gamma}\right)\mathrm{d}\gamma(x,y)\,.$$

With this notation, one has

$$\begin{aligned} &\int_{\Omega^3} c\big(x,\gamma_0(x,y,z),y,\hat{\gamma}(x,y,z)\big)\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\ &= \int_{\rho(y)>0}\left(\int_{\Omega^2} c\big(x,(\gamma_0^{01}|\mu)(x|y),y,(\gamma_1^{01}|\mu)(x|y)\big)\frac{(\gamma_0^{12}|\mu)(y|z)}{\frac{\rho}{\mu}(y)}\mathrm{d}x\,\mathrm{d}z\right)\mathrm{d}\mu(y) \\ &\quad + \int_{\rho(y)=0}\left(\int_{\Omega^2} c\big(x,(\gamma_0^{01}|\mu)(x|y),y,0\big)\delta_y(z)\mathrm{d}x\,\mathrm{d}z\right)\mathrm{d}\mu(y) \\ &= \int_\Omega\left(\int_\Omega c\big(x,(\gamma_0^{01}|\mu)(x|y),y,(\gamma_1^{01}|\mu)(x|y)\big)\mathrm{d}x\right)\mathrm{d}\mu(y) \\ &= J_K(\gamma_0^{01},\gamma_1^{01}) = C_K(\rho_0,\rho_1)\,, \end{aligned}$$

and analogously

$$\int_{\Omega^3} c\big(y,\hat{\gamma}(x,y,z),z,\gamma_1(x,y,z)\big)\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = C_K(\rho_1,\rho_2)\,.$$

One finally obtains

$$
\begin{aligned}
C_K(\rho_0,\rho_2)^{\frac{1}{p}} &\leq \Big(\int_{\Omega^2} c\big(x, \gamma_0^{02}(x,z), z, \gamma_1^{02}(x,z)\big) \mathrm{d}x \, \mathrm{d}z \Big)^{\frac{1}{p}} \\
&\overset{(1)}{\leq} \Big(\int_{\Omega^3} c\big(x, \gamma_0(x,y,z), z, \gamma_1(x,y,z)\big) \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \Big)^{\frac{1}{p}} \\
&\overset{(2)}{\leq} \Big(\int_{\Omega^3} \Big[c\big(x, \gamma_0(x,y,z), y, \hat{\gamma}(x,y,z)\big)^{\frac{1}{p}} + \\
&\qquad\qquad c\big(y, \hat{\gamma}(x,y,z), z, \gamma_1(x,y,z)\big)^{\frac{1}{p}} \Big]^p \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \Big)^{\frac{1}{p}} \\
&\overset{(3}{\leq} \Big(\int_{\Omega^3} c\big(x, \gamma_0(x,y,z), y, \hat{\gamma}(x,y,z)\big) \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \Big)^{\frac{1}{p}} + \\
&\qquad\qquad \Big(\int_{\Omega^3} c\big(y, \hat{\gamma}(x,y,z), z, \gamma_1(x,y,z)\big) \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z \Big)^{\frac{1}{p}} \\
&\overset{(4)}{=} C_K\left(\rho_0,\rho_1\right)^{\frac{1}{p}} + C_K\left(\rho_1,\rho_2\right)^{\frac{1}{p}}
\end{aligned}
$$

where we used (1) the convexity of c, (2) the fact that $c^{1/p}$ satisfies the triangle inequality, (3) Minkowski's inequality and (4) comes from the computations above. Thus $C_K(\cdot,\cdot)^{1/p}$ satisfies the triangle inequality, and is a metric. $\square$

Next, we establish under reasonable assumptions on the cost c that the corresponding unbalanced transport cost C_K is weakly* continuous w.r.t. the marginals. This is important for numerical applications, since it implies robustness when one approximates the original measures with discrete sums of Diracs. It is also crucial to study the well posedness of variational problems involving these metrics, such as for instance the definition of barycenters (see [AC11] for the classical optimal transport case). Moreover it will be a useful tool in our further analysis (e.g. Theorem 4.3 and examples in Sect. 5).

Theorem 3.3. *Assume c satisfies the assumptions of Theorem 3.2 and:*

(A1) Continuity of the spatial metric: $(x,y) \mapsto c(x,1,y,1)$ is continuous.

(A2) Finite cost of mass disappearance: there exists $x \in \Omega$ such that $c(x,1,x,0) < \infty$.

Then C_K is continuous for the weak topology on $\mathcal{M}_+(\Omega)^2$.*

Lemma 3.4. *Under the assumptions of Theorem 3.3, for some $p \in [1,+\infty[$,*

(B1) $(x_0,x_1) \mapsto c(x_0,1,x_1,1)^{1/p}$ is equivalent to the Euclidean metric;

(B2) $x \mapsto c(x,1,x,0)$ is bounded on Ω;

(B3) the family of functions $\{m \mapsto c(x,1,x,m)\}_{x\in\Omega}$ is equicontinuous at $m=1$.

Proof. (B1): Two metrics d_1, d_2 are equivalent when

$$d_1(x_n,x) \to 0 \Leftrightarrow d_2(x_n,x) \to 0\,.$$

Let $d_1(x,y) = |x-y|$ the Euclidean metric on Ω and $d_2(x,y) = c(x,1,y,1)^{1/p}$. Assumption (A1) implies "$\Rightarrow$". Now we want to show that $d_2(x_n,x) \to 0$ implies $d_1(x_n,x) \to 0$. Let $\varepsilon > 0$ and let $S = \Omega \cap B(x,\varepsilon)$ ($B(x,\varepsilon)$ denoting the open ball of radius ε around x). Then $\Omega \setminus S$ is compact. Therefore $x' \mapsto d_2(x,x')$ attains its minimum δ over $\Omega \setminus S$ on that set and by the metric property $\delta > 0$. Since $d_2(x_n,x) \to 0$ there is some $N \in \mathbb{N}$ such that $d_2(x_n,x) < \delta$ for $n > N$ and consequently $x_n \in B(x,\varepsilon)$ for $n > N$.

(B2) For any $y \in \Omega$ one has by the triangle inequality and the identification of zero-mass points $c(y,1,y,0)^{1/p} \le c(y,1,x,1)^{1/p} + c(x,1,x,0)^{1/p}$. The first term is bounded since it is continuous in y and Ω is compact. The second term is bounded by assumption (A2).

(B3) Let C be the uniform bound of $c(x,1,x,0)$ over Ω. By 1-homogeneity and triangle inequality it holds $c(x,1,x,2) \le \left(c(x,1,x,0)^{1/p} + c(x,2,x,0)^{1/p}\right)^p \le C\left(1+2^{1/p}\right)^p \stackrel{\text{def.}}{=} C'$. By convexity of c, $c(x,m,x,1) \le |m-1|\,C'$ for $m \in [0,2]$. As c is nonnegative, the family $\{m \mapsto c(x,1,x,m)\}_{x\in\Omega}$ is therefore uniformly continuous at $m = 1$. ◻

Proof of Theorem 3.3. Throughout this proof we are going to use the properties (B1) to (B3) established in Lemma 3.4. Thanks to the triangle inequality, we only need to check that, if $\rho_n \rightharpoonup^* \rho$ then $C_K(\rho_n,\rho) \to 0$.

Recall that weak* convergence implies (1) convergence of the total masses and (2) convergence for the p-Wasserstein distance W_p ($p \in [1,+\infty[$) of the rescaled measures (see for instance [Vil03, Theorem 7.12]). We denote by $W_{p,c}$ the Wasserstein metric induced by the ground metric $(x,y) \mapsto c(x,1,y,1)^{1/p}$. By property (B1) this ground metric induces the same topology on Ω as the Euclidean metric. Therefore both metrics induce the same space of continuous functions $C(\Omega)$ and therefore convergence of W_p and $W_{p,c}$ is equivalent, too.

If $\rho = 0$ then, by taking $(\gamma_n \stackrel{\text{def.}}{=} (\mathrm{id},\mathrm{id})_\#\rho_n, 0) \in \Gamma(\rho_n,\rho)$ we see that, by property (B2),

$$C_K(\rho_n,\rho) \le \int_{\Omega^2} c(x_0,1,x_1,0)\,\mathrm{d}\gamma_n(x_0,x_1) \le (\sup_{x\in\Omega}|c(x,1,x,0)|)\cdot\rho_n(\Omega) \to 0\,.$$

Otherwise, let us assume that for all $n \in \mathbb{N}$, $\rho_n(\Omega) > 0$ as it is eventually true. Introduce $\mathcal{N}(\rho_n) \stackrel{\text{def.}}{=} (\rho(\Omega)/\rho_n(\Omega))\cdot\rho_n$. It holds, by the triangle inequality,

$$C_K^{\frac{1}{p}}(\rho_n,\rho) \le C_K^{\frac{1}{p}}(\rho_n,\mathcal{N}(\rho_n)) + C_K^{\frac{1}{p}}(\mathcal{N}(\rho_n),\rho)\,.$$

On the one hand, by choosing $\gamma_0 = (\mathrm{id},\mathrm{id})_\#\rho_n$ and $\gamma_1 = (\rho(\Omega)/\rho_n(\Omega))\cdot\gamma_0$ we have, by property (B3),

$$C_K(\rho_n,\mathcal{N}(\rho_n)) \le J_K(\gamma_0,\gamma_1) = \int_\Omega c\big(x,1,x,\frac{\rho(\Omega)}{\rho_n(\Omega}\big)\mathrm{d}\rho_n(x) \to 0\,.$$

On the other hand, by choosing $\gamma_0 = \gamma_1$ equal to the optimal transport plan for the $W_{p,c}$ metric between ρ and $\mathcal{N}(\rho_n)$, we have by property (B1),

$$C_K(\mathcal{N}(\rho_n), \rho) \leq \int_{\Omega^2} c(x, 1, y, 1) \mathrm{d}\gamma_0 \to 0$$

and we conclude thus that $C_K(\rho_n, \rho) \to 0$. □

Let us now give a dual formulation of the problem. Since the cost function c is jointly 1-homogeneous, convex, and l.s.c. in the variables (m_0, m_1), for all $(x, y) \in \Omega^2$, the Legendre transform of $c(x, \cdot, y, \cdot)$ is the indicator of a closed convex set. Morevover, as c is non-negative, and $+\infty$ if m_0 or m_1 is negative, the latter contains the negative orthant. This set is described by the set valued function $Q : \Omega^2 \to 2^{\mathbb{R}\times\mathbb{R}}$. For nonempty convex valued multifunctions, Q is said to be *lower semicontinuous* if $G = \{(x, t) : x \in \operatorname{int} Q(t)\}$ is open (see [Roc71, Lemma 2]).

Theorem 3.5 (Duality)**.** *Let c be a* cost function, *define for all $(x, y) \in \Omega^2$ the nonempty, closed and convex set $Q(x, y) = \{(a, b) \in \mathbb{R}^2 : c(x, \cdot, y, \cdot)^*(a, b) < +\infty\}$ and let*

$$B = \left\{(\phi, \psi) \in C(\Omega)^2 : \forall (x, y) \in \Omega^2, (\phi(x), \psi(y)) \in Q(x, y)\right\} .$$

If Q is lower semicontinuous then

$$\sup_{(\phi,\psi)\in B} \int_\Omega \phi(x) \mathrm{d}\rho_0 + \int_\Omega \psi(y) \mathrm{d}\rho_1 = \min_{(\gamma_0,\gamma_1)\in\Gamma(\rho_0,\rho_1)} J_K(\gamma_0, \gamma_1) .$$

Proof. Let us rewrite the supremum problem as

$$\sup_{(u_0,u_1)\in(C(\Omega^2)^2} -F(u_0, u_1) - G(u_0, u_1)$$

where

$$G : (u_0, u_1) \mapsto \begin{cases} -\int_\Omega \phi(x)\mathrm{d}\rho_0 - \int_\Omega \psi(y)\mathrm{d}\rho_1 & \text{if } u_0(x, y) = \phi(x) \\ & \text{and } u_1(x, y) = \psi(y) \\ +\infty & \text{otherwise,} \end{cases}$$

and F is the indicator of $\{(u_0, u_1) \in (C(\Omega^2)^2) : (u_0, u_1)(x, y) \in Q(x, y), \forall (x, y) \in \Omega^2\}$. Note that F and G are convex and proper. Also, given our assumptions, there is a pair of functions (u_0, u_1) at which F is continuous (for the sup norm topology) and F and G are finite since for all $(x, y) \in \Omega^2$, $Q(x, y)$ contains the negative orthant. Then Fenchel-Rockafellar duality theorem (see, e.g. [Vil03, Theorem 1.9]) states that

$$\sup_{(u_0,u_1)\in(C(\Omega^2)^2} -F(u_0, u_1) - G(u_0, u_1) = \min_{(\gamma_0,\gamma_1)\in\mathcal{M}(\Omega^2)^2} \{F^*(\gamma_0, \gamma_1) + G^*(-\gamma_0, -\gamma_1)\} . \tag{3.5}$$

Let us compute the Legendre transforms. For G, we obtain

$$G^*(-\gamma_0,-\gamma_1) = \sup_{(\phi,\psi)\in C(\Omega)^2} -\int_{\Omega^2}\phi(x)\mathrm{d}\gamma_0 - \int_{\Omega^2}\psi(x)\mathrm{d}\gamma_1 + \int_\Omega \phi(x)\mathrm{d}\rho_0 + \int_\Omega \psi(y)\mathrm{d}\rho_1$$
$$= \begin{cases} 0 & \text{if } (\gamma_0,\gamma_1)\in \Gamma^{+/-}_{\rho_0,\rho_1} \\ +\infty & \text{otherwise.}\end{cases}$$

where $\Gamma^{+/-}_{\rho_0,\rho_1}$ is the set of semi-couplings without the non-negativity constraint. On the other hand, F^* is given by [Roc71, Theorem 6] which states that

$$F^*(\gamma_0,\gamma_1) = \int_{\Omega^2} c\left(x, \tfrac{\gamma_0}{\gamma}, y, \tfrac{\gamma_1}{\gamma}\right)\mathrm{d}\gamma(x,y)$$

where γ is any measure in $\mathcal{M}_+(\Omega^2)$ with respect to which (γ_0,γ_1) is absolutely continuous. Finally, as F^* includes the nonnegativity constraint, the right hand side of (3.5) is equal to $\min_{(\gamma_0,\gamma_1)\in\Gamma(\rho_0,\rho_1)} J_K(\gamma_0,\gamma_1)$. □

4 From Dynamic Models to Static Problems

In this section we establish that a certain class of convex, positively homogeneous, optimization problems over the solutions of the continuity equation with source (informally introduced in 1.2) — the dynamic problems — admits unbalanced Kantorovich formulations that we introduced in Sect. 3 — the static problems. We prove an equivalence result between static and dynamic models for a large class of dynamic models.

4.1 A Family of Dynamic Problems

Dynamic formulations of unbalanced transportation metrics correspond intuitively to the computation of geodesic distances according to a function measuring the infinitesimal effort needed for "acting" on a mass m at position x according to the speed v and rate of growth α (cf. (1.4)). This should be contrasted with the static formulation (3.3) that depends on a cost function $c(x_0,m_0,x_1,m_1)$ between pairs of positions and masses.

The continuity equation, informally introduced in (1.2), is a concept underlying all dynamical formulations of this article. It enforces a local mass preservation constraint for a density ρ, a flow field v and a growth rate field α. We now give a rigorous definition in terms of measures (ρ,ω,ζ) where ω can informally be interpreted as momentum $\rho\cdot v$ of the flow field and ζ corresponds $\rho\cdot\alpha$.

Definition 4.1 (Continuity equation with source)**.** For $(a,b)\in\mathbb{R}^2$ and $\Omega\subset\mathbb{R}^d$ compact, denote by $\mathcal{CE}_a^b(\rho_0,\rho_1)$ the affine subset of $\mathcal{M}([a,b]\times\Omega)\times\mathcal{M}([a,b]\times\Omega)^d\times\mathcal{M}([a,b]\times\Omega)$ of triplets of measures $\mu=(\rho,\omega,\zeta)$ satisfying the continuity equation $\partial_t\rho+\nabla\cdot\omega=\zeta$ in the distribution sense, interpolating between ρ_0 and ρ_1 and satisfying homogeneous Neumann boundary conditions. More precisely we require

$$\int_a^b\int_\Omega \partial_t\varphi\,\mathrm{d}\rho + \int_a^b\int_\Omega \nabla\varphi\cdot\mathrm{d}\omega + \int_a^b\int_\Omega \varphi\,\mathrm{d}\zeta = \int_\Omega \varphi(b,\cdot)\mathrm{d}\rho_1 - \int_\Omega \varphi(a,\cdot)\mathrm{d}\rho_0 \quad (4.1)$$

for all $\varphi \in C^1([a,b]\times\Omega)$.

Below we collect without proof a few standard facts on this equation which will be useful for the following. The notation $B^d(0,r)$ denotes the open ball of radius r in $\mathbb{R}^d$ centered at the origin.

Proposition 4.1.

Glueing *Let $(a,b,c)\in\mathbb{R}^3$ satisfying $a<b<c$, $(\rho_a,\rho_b,\rho_c)\in\mathcal{M}_+(\Omega)^3$, $\mu_A\in\mathcal{CE}_a^b(\rho_a,\rho_b)$ and $\mu_B\in\mathcal{CE}_b^c(\rho_b,\rho_c)$. Then the measure μ defined as μ_A for $t\in[a,b[$ and μ_B for $t\in[b,c]$ belongs to $\mathcal{CE}_a^c(\rho_a,\rho_c)$.*

Smoothing *Let $\varepsilon>0$, let r_ε^x, r_ε^t mollifiers supported on the open balls $B^d(0,\frac{\varepsilon}{2})$ and $B^1(0,\frac{\varepsilon}{2})$ respectively and $r_\varepsilon:(t,x)\mapsto r_\varepsilon^t(t)r_\varepsilon^x(x)$. Let $\mu=(\rho,\omega,\zeta)$ be a triplet of measures supported on $\mathbb{R}\times\Omega$ such that $\mu\in\mathcal{CE}_0^1(\rho_0,\rho_1)$, $\mu=(\rho_0,0,0)\otimes\mathrm{d}t$ for $t<0$, and $\mu=(\rho_1,0,0)\otimes\mathrm{d}t$ for $t>1$. Then for all $a\le-\varepsilon/2$, and $b\ge 1+\varepsilon/2$, $\mu*r_\varepsilon\in\mathcal{CE}_a^b(\rho_0*r_\varepsilon^x,\rho_1*r_\varepsilon^x)$ on $\Omega+\bar{B}^d(0,\varepsilon/2)$.*

Scaling *Let $\mu=(\rho,\omega,\zeta)\in\mathcal{CE}_a^b(\rho_a,\rho_b)$ with $a<b$ and $T:(t,x)\mapsto(T_t(t),T_x(x))$ be an affine scaling with multiplication factor α and β, respectively. Then $(\alpha\cdot T_\#\rho,\beta\cdot T_\#\omega,T_\#\zeta)\in\mathcal{CE}_{T_t(a)}^{T_t(b)}((T_x)_\#(\rho_a),(T_x)_\#(\rho_b))$ on the domain $T_x(\Omega)$.*

Next, we introduce the admissible class of infinitesimal costs which generalizes the admissible Riemannian inner products defined in Section 2.2. Although a standard Lagrangian cost would be defined as a function of the speed v and the growth rate α, our infinitesimal cost has nicer analytical properties when defined in terms of the momentum ω and the source ζ variables. Indeed, it is then natural to require subadditivity (i.e. we expect that "locally", mass is not encouraged to split) and homogeneity in (ρ,ω,ζ) (thus convexity). Finally, this change of variables allows interesting costs where ω or ζ do not necessarily admit a density w.r.t. ρ (see Section 5.1).

Definition 4.2 (Infinitesimal cost)**.** In the following, an infinitesimal cost is a function $f:\Omega\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}\to\mathbb{R}_+\cup\{+\infty\}$ such that for all $x\in\Omega$, $f(x,\cdot,\cdot,\cdot)$ is convex, positively 1-homogeneous, lower semicontinuous and satisfies

$$f(x,\rho,\omega,\zeta)\begin{cases}=0 & \text{if } (\omega,\zeta)=(0,0) \text{ and } \rho\ge 0\\ >0 & \text{if } |\omega| \text{ or } |\zeta|>0\\ =+\infty & \text{if } \rho<0\,.\end{cases}$$

For Theorem 4.3, we also require (i) that there exists continuous functions $\lambda_i:\Omega\to\mathbb{R}_+^*$, $i\in\{1,\dots,N\}$ such that

$$f(x,\rho,\omega,\zeta)=\sum_{i=1}^N\lambda_i(x)\tilde{f}_i(\rho,\omega,\zeta) \tag{4.2}$$

where each $\tilde{f}_i$ satisfies an integrability condition: there exists $C_i>0$ such that $|\tilde{f}_i(\cdot)|\le C_i\max_x f(x,\cdot)$ and (ii) a non-degeneracy condition : there exists $C>0$ such that $f(x,\rho,\omega,2\zeta)\le C\cdot f(x,\rho,\omega,\zeta)$, for all $(x,\rho,\omega,\zeta)\in\Omega\times\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}$.

Note that, in particular, any admissible Riemannian metric in Section 2 satisfies these conditions (by equation (2.5)). The dynamic formulation is now defined as the computation of a geodesic length for the infinitesimal cost f.

Definition 4.3 (Dynamic problem)**.** For $(\rho, \omega, \zeta) \in \mathcal{M}([0,1] \times \Omega)^{1+d+1}$, let

$$J_D(\rho, \omega, \zeta) \stackrel{\text{def.}}{=} \int_0^1 \int_\Omega f(x, \tfrac{\rho}{\lambda}, \tfrac{\omega}{\lambda}, \tfrac{\zeta}{\lambda})\, \mathrm{d}\lambda(t, x) \tag{4.3}$$

where $\lambda \in \mathcal{M}_+([0,1] \times \Omega)$ is such that $(\rho, \omega, \zeta) \ll \lambda$. Due to 1-homogeneity of f this definition does not depend on the choice of λ. The dynamic problem is, for $\rho_0, \rho_1 \in \mathcal{M}_+(\Omega)$,

$$C_D(\rho_0, \rho_1) \stackrel{\text{def.}}{=} \inf_{(\rho,\omega,\zeta) \in \mathcal{CE}_0^1(\rho_0,\rho_1)} J_D(\rho, \omega, \zeta)\,. \tag{4.4}$$

Similarly to the static case (Theorem 3.5), the dynamic setting also enjoys a dual formulation. For the definition of lower semicontinuity for set valued functions, see the preliminaries before Theorem 3.5.

Proposition 4.2 (Duality)**.** *Let $B(x)$ be the polar set of $f(x, \cdot, \cdot, \cdot)$ for all $x \in \Omega$ and assume it is a lower semicontinuous set-valued function. Then the minimum of (4.4) is attained and it holds*

$$C_D(\rho_0, \rho_1) = \sup_{\varphi \in K} \int_\Omega \varphi(1, \cdot)\mathrm{d}\rho_1 - \int_\Omega \varphi(0, \cdot)\mathrm{d}\rho_0 \tag{4.5}$$

with $K \stackrel{\text{def.}}{=} \left\{\varphi \in C^1([0,1] \times \Omega) : (\partial_t \varphi, \nabla\varphi, \varphi) \in B(x),\ \forall (t,x) \in [0,1] \times \Omega\right\}$.

Proof. Remark, that (4.5) can be written as

$$- \inf_{\varphi \in C^1([0,1]\times\Omega)} F(A\varphi) + G(\varphi)$$

where $A : \varphi \mapsto (\partial_t\varphi, \nabla\varphi, \varphi)$, is a bounded linear operator from $C^1([0,1] \times \Omega)$ to $C([0,1] \times \Omega)^{d+2}$, and $F : (\alpha, \beta, \gamma) \mapsto \int_0^1 \int_\Omega \iota_{B(x)}(\alpha(t,x), \beta(t,x), \gamma(t,x))\mathrm{d}x\mathrm{d}t$, $G : \varphi \mapsto \int_\Omega \varphi(0,\cdot)\mathrm{d}\rho_0 - \int_\Omega \varphi(1,\cdot)\mathrm{d}\rho_1$ are convex, proper and lower-semicontinuous functionals, in particular because for all $x \in \Omega$, the set $B(x)$ is convex, closed and nonempty. Since we assumed that $f(x, \rho, \omega, \zeta) > 0$ if $|\omega| > 0$ or $\zeta > 0$ and f is continuous as a function of x on the compact Ω, one can check that there exists $\varepsilon > 0$ such that $(-\varepsilon, 0, \theta\epsilon/2) \in (\text{ int } \cap_{x\in\Omega} B(x))$ for $\theta \in [-1, 1]$ and thus there exists a function $\varphi : t \mapsto -\varepsilon t + \varepsilon/2$ such that $F(A\varphi) + G(\varphi) < +\infty$ and F continuous at $A\varphi$. Then, by Fenchel-Rockafellar duality, (4.5) is equal to

$$\min_{\mu \in \mathcal{M}([0,1]\times\Omega)^{d+2}} G^*(-A^*\mu) + F^*(\mu)\,.$$

By [Roc71, Theorem 6] and the lower-semicontinuity of $x \mapsto B(x)$, we have $F^* = J_D$, and by direct computations, $G^* \circ (-A^*)$ is the convex indicator of $\mathcal{CE}_0^1(\rho_0, \rho_1)$. ☐

4.2 Connection Between Static and Dynamic Problems

Theorem 4.3 is the main result of this section. It states that if the cost c of the static problem is defined as the dynamic cost between Diracs, then the static and the dynamic formulations coincide. The cost c of the static problem can alternatively be determined as a generalized action on the paths space as measured by an infinitesimal cost f (see Proposition 4.4). In the Riemannian setting this means that c is the squared geodesic distance on $\mathrm{Cone}(\Omega)$, see Sect. 2.2. Of course, there are static costs c that do not arise from a dynamic cost f, just as there are dynamic problems beyond the framework of Definition 4.2 which cannot be cast into a static form.

Definition 4.4. The Dirac-based cost is

$$c_d : (x_0, m_0), (x_1, m_1) \mapsto C_D(m_0 \delta_{x_0}, m_1 \delta_{x_1}) \,. \tag{4.6}$$

If c_d is l.s.c. then it defines a *cost function*.

Definition 4.5. The path-based cost c_s is defined by a minimization over smooth trajectories

$$c_s(x_0, m_0, x_1, m_1) \stackrel{\text{def.}}{=} \inf_{(x(t),m(t))} \int_0^1 f(x(t), m(t), m(t)\, x'(t), m'(t))\, \mathrm{d}t \tag{4.7}$$

for $(x(t), m(t)) \in C^1([0,1], \Omega \times [0, +\infty[)$ such that $(x(i), m(i)) = (x_i, m_i)$ for $i \in \{0, 1\}$. In general, c_s does not define a cost function in the sense of Definition 3.1.

Theorem 4.3. *Let $\Omega \subset \mathbb{R}^d$ be a compact which is star shaped w.r.t. a set of points with nonempty interior, and c be a cost function satisfying $c_d \leq c \leq c_s$. If the associated static problem C_K is weakly* continuous, it holds, for $\rho_0, \rho_1 \in \mathcal{M}_+(\Omega)$, $C_D(\rho_0, \rho_1) = C_K(\rho_0, \rho_1)$ and $c = c_d$.*

Sufficient conditions on c for the weak* continuity of C_K are given in Theorem 3.3. The assumptions on the domain include convex sets but also most star-shaped sets. It is however not our intention to look for the weakest assumptions on the domain for our result to hold. In general, computing c_d directly is not easy : the margin in the choice of c in Theorem 4.3 allows to obtain c_d as a consequence of the Theorem, not as a requirement. A natural choice for c is the convex regularization of the cost on the path space c_s, which is easier to compute than c_d. It can be nicely expressed as the optimal cost on the path space when allowing the Dirac to be split into two chunks:

Proposition 4.4. *The convex regularization of c_s can be expressed as*

$$c : (x_0, m_0), (x_1, m_1) \mapsto \inf_{\substack{m_0^a + m_0^b = m_0 \\ m_1^a + m_1^b = m_1}} c_s(x_0, m_0^a, x_1, m_1^a) + c_s(x_0, m_0^b, x_1, m_1^b) \,. \tag{4.8}$$

It is convex, positively homogeneous in (m_0, m_1), and satisfies $c_d \leq c \leq c_s$.

Proof. The fact that (4.8) is the convex regularization of c_s is a consequence of Carathéodory's Theorem (see [Roc15, Corollary 17.1.6]). It is clear that $c_d \leq c \leq c_s$. ☐

Note that in general, $c_d \neq c_s$, since, as opposed to the dynamic problem C_D, the problem defining c_s is not allowed to split mass. For instance, the cut distance is $\pi/2$ for the *WF* metric (see Section 5.2) while it is π for the Riemannian metric on the cone (see Proposition 2.3). We now proceed to the proof of Theorem 4.3.

Proof of Theorem 4.3. It is clear that $c_d \leq c_s$ because for any candidate path $(x(t), m(t))$ in (4.7), $m(t)\delta_{x(t)} \otimes \mathrm{d}t \in \mathcal{CE}_0^1(m_0\delta_{x_0}, m_1\delta_{x_1})$, and thus the assumption $c_d \leq c \leq c_s$ is not void. The proof is divided into four steps: in Step 1, we show that for marginals which are atomic measures, it holds $C_K \geq C_D$. By integrating characteristics (an argument similar to the original proof of the Benamou-Brenier formula [BB00]), we show in Step 2 that for absolutely continuous marginals, C_K is upper bounded by the dynamic minimization problem restricted to smooth fields. In Step 3, a regularization argument (inspired by [Vil03, Theorem 8.1]) then extends this result to general measures and for the actual problem C_D. This is were the weak* continuity of C_K intervenes. The conclusion, in Step 4, follows by a density argument.

Step 1. Let $\rho_0, \rho_1 \in \mathcal{M}_+^{at}(\Omega)$. For the static problem C_K, there exists a minimizer in the set $\Gamma(\rho_0, \rho_1) \cap \mathcal{M}_+^{at}(\Omega^2)^2$. Indeed, an atom assigned to another atom is represented by an atom in Ω^2, and the same can be forced to hold for an atom assigned to the apex, because for all $x \in \Omega$, $c(x, 1, \cdot, 0)$ attains its minimum in Ω. Let (γ_0, γ_1) be such a minimizer, it can be written $\gamma_k = \sum_{i,j} m_{i,j}^k \delta_{(x_i, x_j)}$, for $k = 0, 1$. Since $c \geq c_d$, it holds

$$
\begin{aligned}
C_K(\rho_0, \rho_1) &= \sum_{i,j} c(x_i, m_{i,j}^0, x_j, m_{i,j}^1) \\
&\geq \sum_{i,j} C_D(m_{i,j}^0 \delta_{x_i}, m_{i,j}^1 \delta_{x_j}) \\
&\geq C_D\big(\sum_{i,j} m_{i,j}^0 \delta_{x_i}, \sum_{i,j} m_{i,j}^1 \delta_{x_i}\big) = C_D(\rho_0, \rho_1)
\end{aligned}
$$

where the last inequality is due to the sub-additivity of C_D, inherited from J_D.

Step 2. Let ρ_0, $\rho_1 \in \mathcal{M}_+^{ac}(\Omega)$ be marginals with positive mass and let

$$
(\rho, \omega, \zeta) \in \left\{ (\rho, \omega, \zeta) \in \mathcal{CE}_0^1(\rho_0, \rho_1) : \tfrac{\omega}{\rho}, \tfrac{\zeta}{\rho} \in C^1([0,1] \times \Omega) \right\} .
$$

One can take the Lagrangian coordinates $(\varphi_t(x), \lambda_t(x))$ which are given by the flow of $(v \stackrel{\text{def.}}{=} \omega/\rho, \alpha \stackrel{\text{def.}}{=} \zeta/\rho)$ defined as in Proposition 2.7:

$$
\partial_t \varphi_t(x) = v_t(\varphi_t(x)) \quad \text{and} \quad \partial_t \lambda_t(x) = \alpha_t(\varphi_t(x))\lambda_t(x) \, ,
$$

with the initial condition $(\varphi_0(x), \lambda_0(x)) = (x, 1)$. Recall that $(\varphi_t(x), \lambda_t(x))$ describes the position and the relative increase of mass at time t of a particle initially at position x and that one has $\rho_t = (\varphi_t)_*(\lambda_t \cdot \rho_0)$. It follows,

$$
\begin{aligned}
J_D(\rho,\omega,\zeta) &= \int_{[0,1]\times\Omega} f\big(x,1,v_t(x),\alpha_t(x)\big)\mathrm{d}[(\varphi_t)_*\big(\lambda_t\rho_0\big)](x)\mathrm{d}t \\
&\overset{(1)}{=} \int_\Omega \left[\int_0^1 f\big(\varphi_t(x),1,\partial_t\varphi_t(x),\partial_t\lambda_t(x)/\lambda_t(x)\big)\lambda_t(x)\mathrm{d}t\right]\mathrm{d}\rho_0(x) \\
&\overset{(2)}{=} \int_\Omega \left[\int_0^1 f\big(\varphi_t(x),\lambda_t(x),\lambda_t(x)(\partial_t\varphi_t(x)),\partial_t\lambda_t(x)\big)\mathrm{d}t\right] d\rho_0(x) \\
&\overset{(3)}{\geq} \int_\Omega c(x,1,\varphi_1(x),\lambda_1(x))d\rho_0(x) \\
&\overset{(4)}{\geq} C_K(\rho_0,\rho_1)
\end{aligned}
$$

where we used (1) the change of variables formula (2) homogeneity of f (3) the assumption $c \leq c_s$ and (4) the fact that $\big((\mathrm{id},\varphi_1)_\#\rho_0, (\mathrm{id},\varphi_1)_\#(\lambda_1\rho_0)\big) \in \Gamma(\rho_0,\rho_1)$.

Step 3. Let $\rho_0, \rho_1 \in \mathcal{M}_+(\Omega)$. We want to show, with the help of Step 2, that $C_K(\rho_0,\rho_1) \leq C_D(\rho_0,\rho_1)$. Let $(\rho,\omega,\zeta) \in \mathcal{CE}_0^1(\rho_0,\rho_1)$ and for $\delta \in]0,1[$ let

$$
\tilde{\rho}^\delta = (1-\delta)\rho + \delta\,(\mathrm{d}x\otimes\mathrm{d}t), \quad \tilde{\omega}^\delta = (1-\delta)\omega, \quad \tilde{\zeta}^\delta = (1-\delta)\zeta
$$

so that $\tilde{\rho}^\delta$ is always positive on sets with nonempty interior and $(\tilde{\rho}^\delta, \tilde{\omega}^\delta, \tilde{\zeta}^\delta) \in \mathcal{CE}_0^1(\tilde{\rho}_0^\delta, \tilde{\rho}_1^\delta)$. By convexity,

$$
J_D(\tilde{\rho}^\delta, \tilde{\omega}^\delta, \tilde{\zeta}^\delta) \leq J_D(\rho,\omega,\zeta)\,.
$$

Since $\tilde{\rho}_0^\delta \rightharpoonup^* \rho_0$ and $\tilde{\rho}_1^\delta \rightharpoonup^* \rho_1$ as $\delta \to 0$ and C_K is continuous for the weak* topology, it is sufficient to prove $J_D(\tilde{\rho}^\delta, \tilde{\omega}^\delta, \tilde{\zeta}^\delta) \geq C_K(\tilde{\rho}_0^\delta, \tilde{\rho}_1^\delta)$ for proving $J_D(\rho,\omega,\zeta) \geq C_K(\rho_0,\rho_1)$.

In order to alleviate notations we shall now denote $\tilde{\rho}^\delta, \tilde{\omega}^\delta, \tilde{\zeta}^\delta$ by just ρ,ω,ζ. Also, we denote by $B^d(a,r)$ the open ball of radius r centered at point a in $\mathbb{R}^d$. Up to a translation, we can assume that 0 is in the interior of the set of points w.r.t. which Ω is star shaped. Then [**?**, Theorem 5.3] tells us that the Minkowski gauge $x \mapsto \inf\{\lambda > 0 : x \in \lambda\Omega\}$ is Lipschitz; let us denote by $k \in \mathbb{R}_+^*$ its Lipschitz constant. We introduce the regularizing kernel $r_\varepsilon(t,x) = \frac{1}{\varepsilon^d} r_1\left(\frac{x}{\varepsilon}\right)\frac{1}{\varepsilon} r_2\left(\frac{t}{\varepsilon}\right)$ where $r_1 \in C_c^\infty(B^d(0,\frac{1}{2k}))$, $r_2 \in C_c^\infty(B^1(0,\frac{1}{2k}))$, $r_i \geq 0$, $\int r_i = 1$, r_i even $(i=1,2)$. Let $\bar{\mu} = ((1+\varepsilon)^{-1}\bar{\rho}, (1+\varepsilon)^{-1}\omega, \zeta)$ where $\bar{\rho}$ is a measure on $[-\varepsilon, 1+\varepsilon]\times\Omega$ which is worth ρ on $[0,1]\times\Omega$, $\rho_0\otimes\mathrm{d}t$ on $[-\varepsilon,0[\times\Omega$ and $\rho_1\otimes\mathrm{d}t$ on $]1,1+\varepsilon]\times\Omega$. By the glueing property in Proposition 4.1, $(\bar{\rho},\omega,\zeta)$ still satisfies the continuity equation. Then define

$$
\mu^\varepsilon \overset{\text{def.}}{=} T_\#(\bar{\mu} * r_\varepsilon)|_{[0,1]\times\Omega},
$$

where $T : (t,x) \mapsto ((1+\varepsilon)^{-1}(t+\varepsilon/2), (1+\varepsilon)^{-1}x)$ is built in such a way that the image of the time segment $[-\varepsilon/2, 1+\varepsilon/2]$ is $[0,1]$. Furthermore, since the Minkowski gauge of Ω is k-Lipschitz, the image of $\Omega^\varepsilon \overset{\text{def.}}{=} \Omega + B^d(0,\varepsilon/k)$ by T is

included in Ω. Now, by the smoothing and scaling properties in Proposition 4.1, it holds $\mu^\varepsilon \in \mathcal{CE}_0^1(\rho_0^\varepsilon, \rho_1^\varepsilon)$, in particular because we took care of multiplying ρ and ω by a factor $(1+\varepsilon)^{-1}$ in the definition of $\bar\mu$. Notice that $\rho_0^\varepsilon \rightharpoonup^* \rho_0$ and $\rho_1^\varepsilon \rightharpoonup^* \rho_1$ when $\varepsilon \to 0$ since they are the evaluations of $\bar\mu * r_\varepsilon$ at time $-\varepsilon/2$ and $1+\varepsilon/2$, respectively (also contracted in space by a factor $1+\varepsilon$). Moreover, the vector fields $\omega_t^\varepsilon/\rho_t^\varepsilon$ and $\zeta_t^\varepsilon/\rho_t^\varepsilon$ are well-defined, smooth, bounded functions on $[0,1]\times\Omega$. Therefore, by Step 2,

$$J_D(\mu^\varepsilon) \geq C_K(\rho_0^\varepsilon, \rho_1^\varepsilon).$$

On the other hand, for any $\varepsilon' > 0$, one has

$$\begin{aligned}
J_D(\mu^\varepsilon) &= \int_0^1 \int_\Omega f(y, \tfrac{\mu^\varepsilon}{|\mu^\varepsilon|}) \mathrm{d}|\mu^\varepsilon|(s,y) \\
&\overset{(1)}{=} \int_{-\varepsilon/2}^{1+\varepsilon/2} \int_{\Omega^\varepsilon} f((1-\varepsilon)y, \tfrac{\bar\mu * r_\varepsilon}{|\bar\mu * r_\varepsilon|}) \mathrm{d}|\bar\mu * r_\varepsilon|(s,y) \\
&\overset{(2)}{\leq} \int_{-\varepsilon/2}^{1+\varepsilon/2} \int_{\Omega^\varepsilon} \mathrm{d}|\bar\mu|(t,x) \int_{\mathbb{R}^{1+d}} \mathrm{d}s\mathrm{d}y f((1-\varepsilon)y, \tfrac{\bar\mu}{|\bar\mu|}(t,x)) \cdot r_\varepsilon(s-t, y-x) \\
&\overset{(3)}{=} \sum_i \int_0^1 \int_\Omega \mathrm{d}|\bar\mu|(t,x) \int_{\mathbb{R}^{1+d}} \mathrm{d}s\mathrm{d}y \tilde f_i(\tfrac{\bar\mu}{|\bar\mu|}(t,x)) \lambda_i((1-\varepsilon)y) r_\varepsilon(s-t, y-x) \\
&\overset{(4)}{\leq} \int_0^1 \int_\Omega (1+\varepsilon') f(x, \tfrac{\bar\mu}{|\bar\mu|}) \mathrm{d}|\bar\mu|(t,x)
\end{aligned}$$

where were used (1) the change of variable formula, (2) the convexity and homogeneity of f, (3) the multiplicative dependance in x (with the integrability conditions) assumed on f and (4) the continuity of the strictly positive factors $(\lambda_i)_i$, where ε, chosen small enough, depends on ε'. Therefore $J_D(\bar\mu) \geq J_D(\mu^\varepsilon)(1+\varepsilon')^{-1}$. But by convexity and homogeneity, $J_D(\bar\mu) \leq (1+\varepsilon)^{-1}((1-\varepsilon)J_D(\rho,\omega,\zeta) + \varepsilon J_D(\rho,\omega,2\zeta))$. The term $J_D(\rho,\omega,2\zeta)$ is finite if $J_D(\rho,\omega,\zeta)$ is finite (by our assumption on f) and one has

$$C_K(\rho_0^\varepsilon, \rho_1^\varepsilon) \leq \frac{1+\varepsilon'}{1+\varepsilon}((1-\varepsilon)J_D(\rho,\omega,\zeta) + \varepsilon J_D(\rho,\omega,2\zeta)).$$

Letting ε' and ε go to 0, using the continuity of C_K under weak* convergence and taking the infimum, one recovers in the end $C_K(\rho_0,\rho_1) \leq C_D(\rho,\omega,\zeta)$ as desired.

Step 4. We have proven in Step 1 that if $\rho_0, \rho_1 \in \mathcal{M}_+^{at}(\Omega)$, then

$$C_D(\rho_0,\rho_1) \leq C_K(\rho_0,\rho_1).$$

Moreover, by Step 3, for all $\rho_0,\rho_1 \in \mathcal{M}_+(\Omega)$ on has

$$C_D(\rho_0,\rho_1) \geq C_K(\rho_0,\rho_1).$$

and thus $C_D = C_K$ for atomic measures. But C_D is weakly* l.s.c since J_D is l.s.c. by Reshetnyak lower-semicontinuity (which requires to integrate on an open set, but one can bring ourselves back to that case as in Proposition 3.1). Finally, by density of $\mathcal{M}_+^{at}(\Omega)$ in $\mathcal{M}_+(\Omega)$, for $\rho_0,\rho_1 \in \mathcal{M}_+(\Omega)$, $C_D(\rho_0,\rho_1) \leq C_K(\rho_0,\rho_1)$. The equality $c = c_d$ is direct by computing C_K between Dirac measures, and because c is subadditive. ☐

5 Examples

In this section we discuss some examples which fit into the framework developed in sections 3 and 4. Optimal partial transport is first treated and then our initial motivating example: the Wasserstein-Fisher-Rao metric. In this section, Ω is a convex compact set in $\mathbb{R}^d$.

5.1 An Optimal Partial Transport Problem

We consider an optimal transport problem with relaxed marginal constraints, which for $\rho_0, \rho_1 \in \mathcal{M}_+(\Omega)$ and $p \in \mathbb{N}$, $p \geq 2$, consists in solving

$$\inf_{\tilde{\rho}_0, \tilde{\rho}_1} \frac{1}{p} W_p^p(\tilde{\rho}_0, \tilde{\rho}_1) + \delta \cdot (|\rho_0 - \tilde{\rho}_0|_{TV} + |\rho_1 - \tilde{\rho}_1|_{TV}) \,. \tag{5.1}$$

Proposition 5.1. *The value of the infimum is left unchanged when adding the constraints $\tilde{\rho}_i \leq \rho_i$ ($i = 0, 1$).*

Proof. Let $\gamma \in \mathcal{M}_+(\Omega^2)$ be any coupling between $\tilde{\rho}_0$ and $\tilde{\rho}_1$ and let $\gamma^* \in \mathcal{M}_+(\Omega^2)$ be such that $\gamma^* \leq \gamma$ and $(\mathrm{Proj}_0)_\# \gamma^* = \rho_0 \wedge (\mathrm{Proj}_0)_\# \gamma$. By construction, one has

$$\begin{aligned} |\rho_0 - (\mathrm{Proj}_0)_\#\gamma| - |\rho_0 - (\mathrm{Proj}_0)_\#\gamma^*| &= |(\mathrm{Proj}_0)_\#\gamma - (\mathrm{Proj}_0)_\#\gamma^*| = |\gamma - \gamma^*| \,, \\ |\rho_1 - (\mathrm{Proj}_1)_\#\gamma| - |\rho_1 - (\mathrm{Proj}_1)_\#\gamma^*| &\geq -|(\mathrm{Proj}_1)_\#\gamma - (\mathrm{Proj}_1)_\#\gamma^*| = -|\gamma - \gamma^*| . \end{aligned}$$

By denoting F the functional in (5.1) written as a function of a coupling, it holds

$$F(\gamma) - F(\gamma^*) \geq \int_{\Omega^2} (|y - x|^p / p) \mathrm{d}(\gamma - \gamma^*) \geq 0 \,.$$

A similar truncation procedure for the other marginal leads to the result. ∎

Problem (5.1) is similar to the distances introduced in [PR13] and [PR14] defined as the p-th root of

$$\inf_{\tilde{\rho}_0, \tilde{\rho}_1} W_p(\tilde{\rho}_0, \tilde{\rho}_1) + \delta(|\rho_0 - \tilde{\rho}_0|_{TV} + |\rho_1 - \tilde{\rho}_1|_{TV}) \tag{5.2}$$

with the difference that the problem we consider is invariant under mass rescaling. The link between this problem and the optimal partial transport problem, i.e. an optimal transport problem where one chooses the amount of mass which is transported, was mentioned in [CSPV15] and is recalled below. Note that what follows is also true for to the case $p = 1$ and more general costs on Ω^2, up to a slight adaptation of the duality formulas. The next proposition states that problem (5.1) fits into our framework if we define the cost function as

$$c(x_0, m_0, x_1, m_1) = \min\left(\frac{|x_1 - x_0|^p}{p}, 2\delta\right) \cdot \min(m_0, m_1) + \delta |m_1 - m_0| \tag{5.3}$$

which is l.s.c. and jointly sublinear in the variables (m_0, m_1).

Proposition 5.2 (Link to our framework and optimal partial transport)**.** *Let C_K be the static cost defined as in (3.3), with the cost function (5.3). Then it is equal to (5.1) and one has*

$$C_K(\rho_0,\rho_1) - \delta(\rho_0(\Omega)+\rho_1(\Omega)) = \inf_{\gamma\in\Gamma_{\leq}(\rho_0,\rho_1)} \int_{\Omega^2} (|x_1-x_0|^p/p - 2\delta)\,\mathrm{d}\gamma \tag{5.4}$$

where $\Gamma_{\leq}(\rho_0,\rho_1)$ is the subset of $\mathcal{M}_+(\Omega^2)$ such that the first and second marginals are upper bounded by ρ_0 and ρ_1, respectively.

Remark. The right-hand side of (5.4) is the Lagrangian formulation of the optimal partial transport problem. In this formulation, one replaces the amount of mass m to be transported by a Lagrangian multiplier, which corresponds to 2δ. It is clear that our problem computes an optimal partial transport for some m (the total mass of the optimal γ) but in general, all values of m cannot be recovered by making δ vary (think of atomic measures). This is however the case under the assumptions of [CM10, Corollary 2.11].

Proof. Let us denote by C_{par} the infimum in the right-hand side of (5.4). For $\rho_0,\rho_1\in\mathcal{M}_+(\Omega)$ and any semi-couplings $(\gamma_0,\gamma_1)\in\Gamma(\rho_0,\rho_1)$, let m_0, m_1 be the densities of γ_0, γ_1 w.r.t. some $\gamma\in\mathcal{M}_+(\Omega^2)$ with γ_0, $\gamma_1 \ll \gamma$. Introduce $\bar\gamma = \min(m_0,m_1)\cdot\gamma|_D$ with $D=\{(x_0,x_1)\in\Omega^2 : |x_1-x_0|^p/p\leq 2\delta\}$. It holds $\bar\gamma\in\Gamma_{\leq}(\rho_0,\rho_1)$ and

$$\begin{aligned} J_K(\gamma_0,\gamma_1) &= \int_{\Omega^2} c(x_0,m_0,x_1,m_1)\mathrm{d}\gamma(x_0,x_1) \\ &= \int_{\Omega^2}\frac{1}{p}|x_1-x_0|^p\mathrm{d}\bar\gamma + \delta\int_{\Omega^2}(\mathrm{d}|\gamma_0-\bar\gamma| + \mathrm{d}|\gamma_1-\bar\gamma|) \\ &= \delta|\rho_0|_{TV} + \delta|\rho_1|_{TV} + \int_{\Omega^2}\left(\frac{1}{p}|x_1-x_0|^p - 2\delta\right)\mathrm{d}\bar\gamma\,. \end{aligned}$$

This implies $C_K-\delta(|\rho_0|_{TV}+|\rho_1|_{TV})\geq C_{par}$. The opposite inequality comes from the remark that the infimum defining C_{par} is unchanged by adding the constraint $\mathrm{supp}(\gamma)\subset D$. Let $\gamma\in\Gamma_{\leq}(\rho_0,\rho_1)$ be supported on D and define for $i\in\{0,1\}$ $\mu_i=\rho_i-(\mathrm{Proj}_i)_\#\gamma\in\mathcal{M}_+(\Omega)$. Let us build a couple of semi-couplings

$$\gamma_i = \gamma + \mathrm{diag}_\#(\mu_0\wedge\mu_1) + \mathrm{diag}_\#(\mu_i - \mu_0\wedge\mu_1),\ i\in\{0,1\},$$

where $\mathrm{diag}: x\mapsto(x,x)$ lifts Ω to the diagonal in Ω^2. Decomposed this way, one obtains directly

$$J_K(\gamma_0,\gamma_1) = \delta|\rho_0|_{TV} + \delta|\rho_1|_{TV} + \int_{\Omega^2}\left(\frac{1}{p}|x_1-x_0|^p - 2\delta\right)\mathrm{d}\gamma\,,$$

Hence $C_K-\delta(|\rho_0|_{TV}+|\rho_1|_{TV})\leq C_{par}$. Finally, one has $C_{par}+\delta(|\rho_0|_{TV}+|\rho_1|_{TV}) = \inf$ (5.1) directly, by applying Proposition 5.1 and rewriting (5.1) as a minimization on a variable $\gamma\in\Gamma_{\leq}(\rho_0,\rho_1)$. $\square$

Now consider the infinitesimal cost

$$f : (\rho, \omega, \zeta) \mapsto \begin{cases} \frac{1}{p}\frac{|\omega|^p}{\rho^{p-1}} + \delta|\zeta| & \text{if } \rho > 0 \\ \delta|\zeta| & \text{if } \rho = |\omega| = 0 \\ +\infty & \text{otherwise .} \end{cases} \tag{5.5}$$

which satisfies the conditions of Definition 4.2 (it is implicitly independent of the space variable x) .

Proposition 5.3 (Cost on the path space)**.** *The static and dynamic costs defined in* (5.3) *and* (5.5) *are related by*

$$c(x_0, m_0, x_1, m_1) = \inf_{(x(t),m(t))} \int_0^1 f(m(t), m(t)x'(t), m'(t))\, \mathrm{d}t \tag{5.6}$$

for $(x(\cdot), m(\cdot)) \in C^1([0,1], \Omega \times [0, +\infty[)$*,* $x(i) = x_i$ *and* $m(i) = m_i$*, for* $i = 0, 1$*.*

Proof. First, for any C^1 path (x, m) we denote by $\bar{m} = \min_{t\in[0,1]} m(t)$ its minimum mass. It holds

$$\begin{aligned}\int_0^1 f(m(t), m(t)x'(t), m'(t))\mathrm{d}t &= \int_0^1 \frac{1}{p}|x'(t)|^p m(t)\mathrm{d}t + \delta \int_0^1 |m'(t)|\mathrm{d}t \\ &\geq \bar{m} \int_0^1 \frac{1}{p}|x'(t)|^p \mathrm{d}t + \delta(|m_0 - \bar{m}| + |m_1 - \bar{m}|) \\ &\geq c(x_0, m_0, x_1, m_1)\, .\end{aligned}$$

For the opposite inequality, let us build a minimizing sequence. The infimum is clearly left unchanged when one considers piecewise C^1 trajectories. In the case where $|x_0 - x_1|^p/p \leq 2\delta$, we divide the time interval into three segments $[0, \epsilon], [\epsilon, 1 - \epsilon]$ and $[1 - \epsilon, 1]$ and build a piecewise C^1 trajectory by making pure variation of mass (or staying on place) in the first and the third segment, and constant speed transport of the mass $\bar{m} = \min(m_0, m_1)$ during the second segment. This way, we obtain that the right-hand side of (5.6) is upper bounded by

$$|m_1 - m_0| + \lim_{\varepsilon\to 0} \int_\varepsilon^{1-\varepsilon} \frac{1}{p}|x'(t)|^p \bar{m}\mathrm{d}t = c(x_0, m_0, x_1, m_1)\, .$$

In the case $|x_0 - x_1|^p/p \geq 2\delta$, one obtain the same inequality by building a similar path, but by transporting only an amount ε of mass in the second segment. ☐

Theorem 5.4. *The equivalence between the static and the dynamic formulations* $C_D = C_K$ *holds and* $C_K^{1/p}$ *defines a distance on* $\mathcal{M}_+(\Omega)$ *which is continuous under weak* convergence. We have the dual formulation*

$$C_K(\rho_0, \rho_1) = \sup_{(\phi,\psi)\in C(\Omega)^2} \int_\Omega \phi\, \mathrm{d}\rho_0 + \int_\Omega \psi\, \mathrm{d}\rho_1$$

subject to, for all $(x, y) \in \Omega^2$*,* $\phi(x) + \psi(y) \leq \frac{1}{p}|y - x|^p$ *and* $\phi(x), \psi(y) \leq \delta$*. Equivalently,*

$$C_K(\rho_0, \rho_1) = \sup_{\varphi\in C^1([0,1]\times\Omega)} \int_\Omega \varphi(1, \cdot)\mathrm{d}\rho_1 - \int_\Omega \varphi(0, \cdot)\mathrm{d}\rho_0\, ,$$

subject to $|\varphi| \leq \delta$ *and* $\partial_t \varphi + \frac{p-1}{p} |\nabla \varphi|^{\frac{p}{p-1}} \leq 0$.

Proof. We first prove that $c^{\frac{1}{p}}$ defines a metric on Cone(Ω). This is mainly a consequence of the inequality

$$(a+c)^p + b + d \leq \left((a^p + b)^{\frac{1}{p}} + (c^p + d)^{\frac{1}{p}} \right)^p$$

which holds true for $(a,b,c,d) \in \mathbb{R}^4_+$ and $p \in \mathbb{N}^*$ (this becomes clear when the right-hand term is expanded). Thus, if we take $A = (x_0, m_0)$, $B = (x_1, m_1)$ and $C = (x_2, m_2)$, three points in $\Omega \times [0, +\infty[$ satisfying $\frac{1}{p}|x_2 - x_0|^p \leq 2\delta$ (the other case is easy) and, without loss of generality, $1 = m_0 \leq m_2$. Remark that there always exists $\tilde{B} = (\tilde{x}_1, \tilde{m}_1)$ satisfying $|x_0 - \tilde{x}_1|^p \leq 2p\delta$, $|x_2 - \tilde{x}_1|^p \leq 2p\delta$, $m_0 \leq \tilde{m}_1 \leq m_2$ and $c(A,B) \geq c(A,\tilde{B})$, $c(C,B) \geq c(C,\tilde{B})$. We drop the $1/p$ factor for clarity and obtain

$$\begin{aligned} c(A,C) &\leq (|x_2 - \tilde{x}_1| + |\tilde{x}_1 - x_0|)^p + (|m_2 - \tilde{m}_1| + |\tilde{m}_1 - m_0|) \\ &\leq \left((|x_2 - \tilde{x}_1|^p + |m_2 - \tilde{m}_1|)^{\frac{1}{p}} + (|\tilde{x}_1 - x_0|^p + |\tilde{m}_1 - m_0|)^{\frac{1}{p}} \right)^p \\ &\leq \left(c(A,\tilde{B})^{\frac{1}{p}} + c(\tilde{B},C)^{\frac{1}{p}} \right)^p \leq \left(c(A,B)^{\frac{1}{p}} + c(B,C)^{\frac{1}{p}} \right)^p . \end{aligned}$$

Thus c satisfies all conditions of Theorem 3.3. This implies that C_K is continuous in the weak* topology and thus Theorem 4.3 applies. The duality results are consequences of Proposition 4.2, Theorem 3.5 and direct calculations. ☐

5.2 A Static Formulation for *WF*

Definition 5.1 (The *WF* distance [CSPV15, KMV15]). For a parameter $\delta \in]0, +\infty[$ consider the convex, positively homogeneous, l.s.c. function

$$f : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R} \ni (\rho, \omega, \zeta) \mapsto \begin{cases} \frac{1}{2} \frac{|\omega|^2 + \delta^2 \zeta^2}{\rho} & \text{if } \rho > 0\,, \\ 0 & \text{if } (\rho,\omega,\zeta) = (0,0,0)\,, \\ +\infty & \text{otherwise,} \end{cases} \tag{5.7}$$

and define, for $\rho_0, \rho_1 \in \mathcal{M}_+(\Omega)$,

$$WF(\rho_0, \rho_1)^2 \stackrel{\text{def.}}{=} \inf_{(\rho,\omega,\zeta) \in \mathcal{CE}_0^1(\rho_0,\rho_1)} \int_{[0,1]\times\Omega} f\left(\tfrac{\rho}{\lambda}, \tfrac{\omega}{\lambda}, \tfrac{\zeta}{\lambda} \right) \mathrm{d}\lambda \tag{5.8}$$

where $\lambda \in \mathcal{M}_+([0,1]\times\Omega)$ chosen such that $\rho, \omega, \zeta \ll \lambda$. Due to the 1-homogeneity of f the integral does not depend on the choice of λ.

We now show that *WF* admits a static formulation, which belongs to the class of models introduced in Section 3. First, it is clear that *WF* fits into the previous framework if we choose the cost function c to be $WF(m_0\delta_{x_0}, m_1\delta_{x_1})^2$. This distance has been computed in [CSPV15] and is given by

$$c(x_0, m_0, x_1, m_1) = 2\delta^2 \left(m_0 + m_1 - 2\sqrt{m_0\, m_1} \cdot \overline{\cos}(|x_0 - x_1|/(2\delta))\,. \right) \tag{5.9}$$

where $\overline{\cos} : z \mapsto \cos(|z| \wedge \frac{\pi}{2})$. Remark that c is a cost function and that its square root defines a distance on $\Omega \times \mathbb{R}_+$ (where we identify points with zero mass $\Omega \times \{0\}$) since it is derived from the *WF* distance restricted to Dirac measures. It is direct to see that c satisfies the assumptions of Theorem 3.3, for $p = 2$.

Theorem 5.5 (Continuous static formulation)**.** *Choosing the cost function* (5.9)*, it holds*

$$\mathit{WF}^2(\rho_0, \rho_1) = \min_{(\gamma_0,\gamma_1)\in\Gamma(\rho_0,\rho_1)} J_K(\gamma_0, \gamma_1)\,. \tag{5.10}$$

Proof. This is a particular case of Theorem 4.3. □

Remark. This theorem can be reformulated, in a nutshell, as

$$\frac{1}{2\delta^2}\mathit{WF}^2(\rho_0, \rho_1) = |\rho_0|_{TV} + |\rho_1|_{TV} + \inf_{(\gamma_0,\gamma_1)\in\Gamma(\rho_0,\rho_1)} -2\int_{|y-x|<\pi} \cos(|y-x|/(2\delta))\mathrm{d}(\sqrt{\gamma_0\gamma_1})(x, y)\,.$$

where $\sqrt{\gamma_0\gamma_1} \stackrel{\text{def.}}{=} \left(\frac{\gamma_0}{\gamma}\frac{\gamma_1}{\gamma}\right)^{\frac{1}{2}} \gamma$ for any γ such that $\gamma_0, \gamma_1 \ll \gamma$.

Corollary 5.6. *It holds*

$$\begin{aligned}
\frac{1}{2\delta^2}\mathit{WF}^2(\rho_0, \rho_1) \quad &= \sup_{(\phi,\psi)\in C(\Omega)^2} \int_\Omega \phi(x)\mathrm{d}\rho_0 + \int_\Omega \psi(y)\mathrm{d}\rho_1 \\
\text{subject to, } \forall (x, y) \in \Omega^2\text{:} \quad & \phi(x) \le 1\,, \quad \psi(y) \le 1\,, \\
& (1 - \phi(x))(1 - \psi(y)) \ge \overline{\cos}^2\left(|x - y|/(2\delta)\right)\,.
\end{aligned}$$

Proof. By direct computations we find that $c(x, \cdot, y, \cdot) = \iota^*_{Q(x,y)}$ with

$$Q(x, y) = \left\{(a, b) \in \mathbb{R}^2 : a, b \le 1 \text{ and } (1 - a)(1 - b) \ge \overline{\cos}^2\left(|y - x|/(2\delta)\right)\right\}$$

and apply Theorem 3.5. □

5.3 Γ-convergence of Static *WF*

In [CSPV15] the limit of the growth penalty parameter $\delta \to \infty$ of *WF* is studied and related to classical optimal transport. Here we give the corresponding result for the static problems in terms of Γ-convergence [Bra02]. Recall that this implies both convergence of the optimal values as well as convergence of minimizers. We now denote by c_δ the cost defined in (5.9) to emphasize its dependency on δ.

Theorem 5.7 ((Almost) Classical OT as Limit of *WF*)**.** *Consider the following two generalized static optimal transport problems:*

$$J_\delta(\gamma_0, \gamma_1) = \int_{\Omega^2} c_\delta\big(x, y, \gamma_0(x, y), \gamma_1(x, y)\big)\,\mathrm{d}x\,\mathrm{d}y - 2\,\delta^2(\sqrt{\gamma_0(\Omega^2)} - \sqrt{\gamma_1(\Omega^2)})^2 \tag{5.11}$$

$$J_\infty(\gamma_0, \gamma_1) = \begin{cases} 0 & \text{if } \gamma_0 = 0 \text{ or } \gamma_1 = 0\,, \\ \int_{\Omega^2} |x - y|^2 \mathrm{d}\gamma_0(x, y) \cdot \frac{\sqrt{\alpha}}{2} & \text{if } \gamma_1 = \alpha\,\gamma_0 \text{ for some } \alpha > 0\,, \\ \infty & \text{otherwise.} \end{cases} \tag{5.12}$$

Then J_δ Γ-converges to J_∞ as $\delta \to \infty$.

Remark. One has $\lim_{\delta\to\infty} \mathit{WF}(\rho_0, \rho_1) = \infty$ if $\rho_0(\Omega) \neq \rho_1(\Omega)$. Consequently, to properly study the limit, we subtract the diverging terms in (5.11). Conversely, we slightly modify the classical OT functional, to assign finite cost when the two couplings are strict multiples of each other. The corresponding optimization problem is solved by computing the optimal transport plan between normalized marginals and then multiplying by the geometric mean of the marginal masses.

The proof uses the following Lemma.

Lemma 5.8 (Sqrt-Measure)**.** *Let $A \subset \mathbb{R}^n$ be a compact set. The function*

$$\mathcal{M}_+(A)^2 \ni \mu \mapsto -\sqrt{\mu_1 \cdot \mu_2}(A) \tag{5.13}$$

is weakly l.s.c. and bounded from below by $\sqrt{\mu_1(A)} \cdot \sqrt{\mu_2(A)}$. The lower bound is only obtained, if $\mu_1 = 0$ or $\mu_2 = 0$ or $\mu_1 = \alpha \cdot \mu_2$ for some $\alpha > 0$.*

Proof. With $f(x) = (\sqrt{x_1} - \sqrt{x_2})^2/2$ and a reference measure $\nu \in \mathcal{M}_+(A)$ with $\mu \ll \nu$ we can write

$$-\sqrt{\mu_1 \cdot \mu_2}(A) = \int_A f\left(\frac{\mu}{\nu}\right) \mathrm{d}\nu - \mu_1(A)/2 - \mu_2(A)/2$$

Since f is 1-homogeneous, the evaluation does not depend on the choice of ν. As f is convex, l.s.c., bounded from below, A is bounded, and total masses converge, lower semi-continuity of the functional now follows from [AFP00, Thm. 2.38] (see proof of Proposition 3.1 for adaption to Ω closed).

For the lower bound, let $\mu_1 = \lambda \cdot \mu_2 + \mu_{1,\perp}$ be the Radon-Nikodým decomposition of μ_1 w.r.t. μ_2. Then have

$$-\sqrt{\mu_1 \cdot \mu_2}(A) = -\int_A \sqrt{\lambda}\, \mathrm{d}\mu_2 \geq -\left(\int_A \lambda\, \mathrm{d}\mu_2 \cdot \mu_2(A)\right)^{1/2} \geq -\left(\mu_1(A) \cdot \mu_2(A)\right)^{1/2}$$

where the first inequality is due to Jensen's inequality, with equality only if λ is constant. The second inequality is only an equality if $\mu_{1,\perp} = 0$. $\square$

Proof of Theorem 5.7. **Lim-Sup.** For every pair (γ_0, γ_1) a recovery sequence is given by the constant sequence $(\gamma_0, \gamma_1)_{n\in\mathbb{N}}$. The cases $\gamma_i = 0$ for $i = 0$ or 1, and $\gamma_1 \neq \alpha\, \gamma_0$ for every $\alpha > 0$ are trivial. Therefore, let now $\gamma_1 = \alpha\, \gamma_0$ for some $\alpha > 0$. We find

$$\begin{aligned} J_\delta(\gamma_0, \alpha\, \gamma_0) = &\int_{\Omega^2} 2\, \delta^2 \left[(1+\alpha) - 2\sqrt{\alpha}\, \overline{\cos}(|x-y|/(2\delta))\right] \mathrm{d}\gamma_0(x,y) \\ &- 2\, \delta^2\, \gamma_0(\Omega^2)\, (1-\sqrt{\alpha})^2 \end{aligned}$$

Now use $\overline{\cos}(z) \geq 1 - z^2/2$ to find:

$$\begin{aligned} &\leq \int_{\Omega^2} 4\, \delta^2\, \sqrt{\alpha} \frac{|x-y|^2}{8\, \delta^2}\, \mathrm{d}\gamma_0(x,y) \\ &= J_\infty(\gamma_0, \gamma_1) \end{aligned}$$

Lim-Inf. For a sequence of couplings $(\gamma_{0,k},\gamma_{1,k})_{k\in\mathbb{N}}$ converging weakly* to some pair $(\gamma_{0,\infty},\gamma_{1,\infty})$ we now study the sequence of values $J_k(\gamma_{0,k},\gamma_{1,k})$. Note first, that J_k is weakly* l.s.c. since the integral part is l.s.c. (cf. Proposition 3.1) and the second term is continuous (total masses converge).

Since Ω is compact, there is some $N_1 \in \mathbb{N}$ such that for $k > N_1$, we have

$$1 - z^2/2 \leq \overline{\cos}(z) \leq 1 - z^2/2 + z^4/24 \qquad \text{for} \qquad z = |x-y|/(2\,k),\ x,y \in \Omega\,.$$

And therefore for $k > N_1$ and any coupling pair, denoting $A \overset{\text{def.}}{=} \sqrt{\gamma_0(\Omega^2)} - \sqrt{\gamma_1(\Omega^2)}$,

$$\begin{aligned}
J_k(\gamma_0,\gamma_1) = 2\,k^2 &\left(\int_{\Omega^2} \left(\sqrt{\gamma_0(x,y)} - \sqrt{\gamma_1(x,y)}\right)^2 \mathrm{d}x\,\mathrm{d}y - A^2 \right) \\
&+ 4\,k^2 \int_{\Omega^2} \frac{|x-y|^2}{8\,k^2} \sqrt{\gamma_0(x,y)\,\gamma_1(x,y)}\,\mathrm{d}x\,\mathrm{d}y \\
&- I \cdot 4\,k^2 \int_{\Omega^2} \frac{|x-y|^4}{24\cdot(2\,k)^4} \sqrt{\gamma_0(x,y)\,\gamma_1(x,y)}\,\mathrm{d}x\,\mathrm{d}y
\end{aligned}$$

for some $I \in [0,1]$.

Since Ω is bounded, by means of Lemma 5.8 and since the total masses of $\gamma_{i,k}$, $i = 0,1$ are converging towards the total masses of $\gamma_{i,\infty}$ as $k \to \infty$, there is a constant $C > 0$ and some $N_2 \geq N_1$ such that the coefficient for I in the third line can be bounded by C/k^2 for $k > N_2$ for when calling with arguments $(\gamma_{0,k},\gamma_{1,k})$.

For the first line we write briefly $2\,k^2\,F(\gamma_0,\gamma_1)$. From Lemma 5.8 we find that F is weakly* l.s.c., $F \geq 0$ and $F(\gamma_0,\gamma_1) = 0$ if and only if $(\gamma_0,\gamma_1) \in \mathcal{S}$ with

$$\mathcal{S} = \left\{ (\gamma_0,\gamma_1) \in \mathcal{M}_+(\Omega^2)^2 : \ \gamma_0 = 0 \text{ or } \gamma_1 = 0 \text{ or } \gamma_1 = \alpha \cdot \gamma_0 \text{ for some } \alpha > 0 \right\}\,.$$

It follows that for $N_2 < k_1 < k_2$ one has

$$J_{k_2}(\gamma_{0,k_2},\gamma_{1,k_2}) = J_{k_1}(\gamma_{0,k_2},\gamma_{1,k_2}) + 2\,(k_2^2 - k_1^2)\,F(\gamma_{0,k_2},\gamma_{1,k_2}) + I \cdot C/k_1^2$$

for some $I \in [-1,1]$. Now consider the joint limit:

$$\begin{aligned}
\liminf_{k\to\infty} J_k(\gamma_{0,k},\gamma_{1,k}) &\geq \liminf_{k\to\infty} J_{k_1}(\gamma_{0,k},\gamma_{1,k}) + \liminf_{k\to\infty} 2(k^2 - k_1^2)\,F(\gamma_{0,k},\gamma_{1,k}) - C/k_1^2 \\
&\geq J_{k_1}(\gamma_{0,\infty},\gamma_{1,\infty}) + 2(k_2^2 - k_1^2)\,F(\gamma_{0,\infty},\gamma_{1,\infty}) - C/k_1^2
\end{aligned}$$

for any $N_2 < k_1 < k_2$ (by using weak* l.s.c. of J_{k_1} and F and non-negativity of F). Since $J_{k_1} > -\infty$ and k_2 can be chosen arbitrarily large, we find

$$\liminf_{k\to\infty} J_k(\gamma_{0,k},\gamma_{1,k}) = \infty \quad \text{for} \quad (\gamma_{0,\infty},\gamma_{1,\infty}) \notin \mathcal{S}\,.$$

By reasoning analogous to the lim-sup case (adding the z^4 term in the $\overline{\cos}$-expansion to get a lower bound and bounding its value as above) we find:

$$\liminf_{k\to\infty} J_k(\gamma_{0,k},\gamma_{1,k}) \geq J_\infty(\gamma_{0,\infty},\gamma_{1,\infty}) \quad \text{for} \quad (\gamma_{0,\infty},\gamma_{1,\infty}) \in \mathcal{S}\,.$$

□

Conclusion and Perspectives

In this paper, we presented a unified treatment of the unbalanced transport that allows for both statical and dynamical formulations. Our key findings are (i) a Riemannian submersion from a semi-direct product of groups with an L^2 metric to the *WF* metric, which leads to the computation of the sectional curvature and a Monge formulation, (ii) a new class of static optimal transport formulations involving semi-couplings, (iii) an equivalence between these static formulations and a class of dynamic formulations. Each of these contributions is of independent interest, but the synergy between the static, the dynamic and the Monge problems allows to get a clear picture of the unbalanced transportation problem. Beside these theoretical advances, we believe that a key aspect of this work is that the proposed static formulation opens the door to a new class of numerical solvers for unbalanced optimal transport. These solvers should leverage the specific structure of the cost c considered for each application, a striking example being the *WF* cost (5.9).

Acknowledgements

The work of Bernhard Schmitzer has been supported by the Fondation Sciences Mathématiques de Paris. The work of Gabriel Peyré has been supported by the European Research Council (ERC project SIGMA-Vision). The work of François-Xavier Vialard has been supported by the CNRS (Défi Imag'in de la Mission pour l'Interdisciplinarité, project CAVALIERI).
We would like to thank Yann Brenier for stimulating discussions as well as Peter Michor, in particular for several important references.

A Wasserstein-Fisher-Rao as a Weak Metric

In this appendix, we show that the Wasserstein-Fisher-Rao is a weak metric in a Sobolev setting but does not admit a Levi-Civita connection. Similar results certainly hold for the Wasserstein metric.

We will work in a smooth Sobolev setting, namely on $\mathrm{Dens}^s(\Omega)$ the space of C^1 positive functions on Ω that are in $H^s(\Omega,\mathbb{R})$ for $s > d/2+1$ where d is the dimension of the ambient space of Ω. It is an open subset of $H^s(\Omega,\mathbb{R})$. Note that the same results probably hold for the Wasserstein metric.

Proposition A.1. *The WF metric is a weak Riemannian metric on* $\mathrm{Dens}^s(\Omega)$.

Proof. We use the fact that $\mathrm{Dens}^s(\Omega)$ is an open subset of the Hilbert space $H^s(\Omega,\mathbb{R})$ to work in this coordinate system. The tangent space of $\mathrm{Dens}^s(\Omega)$ is $\mathrm{Dens}^s(\Omega) \times H^s(\Omega,\mathbb{R})$. Let $X \in H^s(\Omega,\mathbb{R})$ be a function that will be seen as a tangent vector at any density $\rho \in \mathrm{Dens}^s(\Omega)$. We denote by $W\!F(\rho)(X,X)$ the *WF* metric evaluated at the point ρ on the tangent vector X. We have to prove that the map from $\mathrm{Dens}^s(\Omega) \times H^s(\Omega)$ into $\mathbb{R}$ defined by

$$(\rho, X) \mapsto W\!F(\rho)(X,X)$$

is smooth. Recall that, using the formulation (2.12), $\mathit{WF}(\rho)(X,X)$ is given by

$$\mathit{WF}(\rho)(X,X) = \frac{1}{2}\langle L(\rho)^{-1}(X), X\rangle_{L^2(\Omega)} \,. \tag{A.1}$$

where $L(\rho) : H^{s+1}(\Omega) \mapsto H^{s-1}(\Omega)$ is the elliptic operator defined by

$$L(\rho)(\phi) = -\nabla\cdot(\rho\nabla\phi) + \rho\,\phi \,. \tag{A.2}$$

Therefore the smoothness of $\mathit{WF}(\rho)(X,X)$ reduces to the smoothness of $L(\rho)^{-1}$ (defined with homogeneous Neumann boundary conditions) as an operator from H^{s-1} into H^{s+1} with respect to ρ. Since $L(\rho)$ is linear in ρ and using the inverse function theorem on Hilbert manifolds, we get the result for the H^{s-1} topology in the second variable X, which is even stronger than the desired result. □

Following [MMM13], we show the non existence of the Levi-Civita connection for the *WF* metric in the Sobolev setting.

Proposition A.2. *The Levi-Civita connection associated with the WF metric does not exist on* $\mathrm{Dens}^s(\Omega)$.

Proof. From [MMM13, page 8], there exists a Levi-Civita associated with a weak Riemannian metric if and only if the metric itself admits gradients with respect to itself in both variables. Let $(\rho, X) \in \mathrm{Dens}^s(\Omega)\times H^s(\Omega)$ be an element of the tangent space. The differentiation with respect to ρ of $\mathit{WF}(\rho)(X,X)$ gives the following L^2 gradient in the direction $Y \in H^s(\Omega,\mathbb{R})$:

$$\partial_\rho \mathit{WF}(\rho)(X,X)(Y) = \frac{1}{2}\langle |\phi|^2 + |\nabla\phi|^2, Y\rangle_{L^2(\Omega)} \,, \tag{A.3}$$

where $\phi = L(\rho)^{-1}(X)$.

The gradient with respect to the *WF* metric is then defined as $L(\rho)(Z)$ where $Z \overset{\text{def.}}{=} \frac{1}{2}(|\phi|^2 + |\nabla\phi|^2)$. However, $Z \in H^s(\Omega)$ and therefore $L(\rho)(Z) \in H^{s-2}$. Thus, the gradient with respect to ρ does not belong in general to H^s. Thus, the Levi-Civita does not exist. □

Note that the key point lies in the loss of smoothness when applying the elliptic operator. This negative result only means that **in this H^s topology**, the weak Riemannian metric *WF* does not admit a Levi-Civita connection. However, this result does not preclude the existence of a topology for which the Levi-Civita connection exists.

B Proof of Proposition 2.5

Proof. For given positive functions a, λ on Ω, one has

$$\Phi^*(mg + \frac{c(x)}{m}\mathrm{d}m^2) = m(g + c(\mathrm{d}\lambda)^2) + 2c\lambda\mathrm{d}\lambda\mathrm{d}m + \frac{c}{m}\lambda^2\mathrm{d}m^2 \,. \tag{B.1}$$

Using that $h(x,m) = mh(x) + a(x)\mathrm{d}m + b(x)\frac{\mathrm{d}m^2}{m}$, then, the result is satisfied if and only if the following system has a solution

$$\begin{cases} c\,\mathrm{d}(\lambda^2) = a \\ c\lambda^2 = b\,. \end{cases} \tag{B.2}$$

Therefore, since c, λ, b are positive functions, dividing the first equation by the second gives:

$$\frac{\mathrm{d}(\lambda^2)}{\lambda^2} = \frac{a}{b}\,.$$

This equation has a solution if and only if $\frac{a}{b}$ is exact. Then, c can then be deduced using the second equation of the system.

The last point consists in proving that $g \overset{\text{def.}}{=} h - c(\mathrm{d}\lambda)^2$ is a metric on Ω. Using the relation in system (B.2), we get $c(\mathrm{d}\lambda)^2 = \frac{a^2}{4b}$. Let us consider $(v_x, v_m) \in T_{(x,m)}(M \times \mathbb{R}_+^*)$ for a non-zero vector v_x, then $mh(x)(v_x,v_x) + a(x)(v_x)v_m + \frac{b(x)}{m}v_m^2$ is a polynomial function in v_m whose discriminant is necessarily strictly negative since $(v_x, v_m) \neq 0$ for all v_m. Therefore, we obtain

$$a(x)(v_x)^2 < 4b(x)h(x)(v_x, v_x) \tag{B.3}$$

which gives $h(x)(v_x,v_x) - c(x)\mathrm{d}\lambda(x)(v_x)^2 > 0$. $\square$

References

[AC11] M. Agueh and G. Carlier. Barycenters in the Wasserstein space. *SIAM J. on Mathematical Analysis*, 43(2):904–924, 2011.

[AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*, volume 254. Clarendon Press Oxford, 2000.

[AG13] L. Ambrosio and N. Gigli. *A user's guide to optimal transport*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2013.

[BB00] J-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[BBI01] D. Burago, Y. Burago, and S. Ivanov. A course in metric geometry. *American Mathematical Soc.*, 2001.

[Ben03] J-D. Benamou. Numerical resolution of an "unbalanced" mass transport problem. *ESAIM: Mathematical Modelling and Numerical Analysis*, 37(05):851–868, 2003.

[Bes78] A. L. Besse. *Manifolds all of whose geodesics are closed / Arthur L. Besse*. Springer-Verlag Berlin ; New York, 1978.

[Bra02] Andrea Braides. *Gamma-Convergence for Beginners*. Oxford University Press, 2002.

[BV13] M. Bruveris and F-X. Vialard. On groups of diffeomorphisms. *Preprint*, 2013.

[CM10] L. Caffarelli and R. J. McCann. Free boundaries in optimal transport and Monge-Ampere obstacle problems. *Annals of mathematics*, 171(2):673–730, 2010.

[CSPV15] L. Chizat, B. Schmitzer, G. Peyré, and F-X. Vialard. An interpolating distance between optimal transport and Fisher-Rao. http://arxiv.org/abs/1506.06430, 2015.

[Del09] Philippe Delanoë. Differential geometric heuristics for riemannian optimal mass transportation. In Boris Kruglikov, Valentin Lychagin, and Eldar Straume, editors, *Differential Equations - Geometry, Symmetries and Integrability*, volume 5 of *Abel Symposia*, pages 49–73. Springer Berlin Heidelberg, 2009.

[EM70] D. G. Ebin and J. E. Marsden. Groups of diffeomorphisms and the motion of an incompressible fluid. *Ann. of Math*, 92:102–163, 1970.

[FG89] Daniel S. Freed and David Groisser. The basic geometry of the manifold of riemannian metrics and of its quotient by the diffeomorphism group. *Michigan Math. J.*, 36(3):323–344, 1989.

[Fig10] A. Figalli. The optimal partial transport problem. *Archive for rational mechanics and analysis*, 195(2):533–560, 2010.

[FZM+15] C. Frogner, C. Zhang, H. Mobahi, M. Araya-Polo, and T. Poggio. Learning with a wasserstein loss. Preprint 1506.05439, Arxiv, 2015.

[Gal79] S. Gallot. Équations différentielles caractéristiques de la sphère. *Annales scientifiques de l'Ecole Normale Superieure*, 12(2):235–267, 1979.

[GPC15] A. Gramfort, G. Peyré, and M. Cuturi. Fast optimal transport averaging of neuroimaging data. In *Proc. MICCAI'15*, 2015.

[Gui02] K. Guittet. Extended Kantorovich norms: a tool for optimization. Technical report, Tech. Rep. 4402, INRIA, 2002.

[Han99] L.G. Hanin. An extension of the Kantorovich norm. *Contemporary Mathematics*, 226:113–130, 1999.

[KMV15] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal trasnport distance on the space of finite Radon measures. Technical report, Pre-print, 2015.

[KR58] L.V. Kantorovich and G.S. Rubinshtein. On a space of totally additive functions. *Vestn Lening. Univ.*, 13:52–59, 1958.

[KW08] B. Khesin and R. Wendt. *The geometry of infinite-dimensional groups*, volume 51. Springer Science & Business Media, 2008.

[Lan99] Serge Lang. *Fundamentals of differential geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1999.

[LM13] D. Lombardi and E. Maitre. Eulerian models and algorithms for unbalanced optimal transport. <hal-00976501v3>, 2013.

[LMS15a] M. Liero, A. Mielke, and G. Savaré. Optimal Entropy-Transport problems and a new Hellinger-Kantorovich distance between positive measures. *ArXiv e-prints*, August 2015.

[LMS15b] M. Liero, A. Mielke, and G. Savaré. Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves. *ArXiv e-prints*, August 2015.

[Lot08] John Lott. Some geometric calculations on Wasserstein space. *Communications in Mathematical Physics*, 277(2):423–437, 2008.

[Mic08] P. W. Michor. *Topics in Differential Geometry*, volume 93 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2008.

[Mis93] Gerard Misiolek. Stability of flows of ideal fluids and the geometry of the group of diffeomorphisms. *Indiana Univ. Math. J.*, 42(1):215 – 237, 1993.

[MMM13] M. Micheli, P. W. Michor, and D. Mumford. Sobolev metrics on diffeomorphism groups and the derived geometry of spaces of submanifolds. *Izvestiya: Mathematics*, 77(3):541, 2013.

[MRSS15] J. Maas, M. Rumpf, C. Schönlieb, and S. Simon. A generalized model for optimal transport of images including dissipation and density modulation. arXiv:1504.01988, 2015.

[Omo78] Hideki Omori. On Banach-Lie groups acting on finite dimensional manifolds. *Tôhoku Math. J.*, 30(2):223–250, 1978.

[O'R97] Donald. O'Regan. *Existence Theory for Nonlinear Ordinary Differential Equations*. Springer, 1997.

[Ott01] F. Otto. The geometry of dissipative evolution equations: The porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, 2001.

[PR13] B. Piccoli and F. Rossi. On properties of the Generalized Wasserstein distance. arXiv:1304.7014, 2013.

[PR14] B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211(1):335–358, 2014.

[PW08] O. Pele and M. Werman. A linear time histogram metric for improved sift matching. In *European Conference on Computer Vision, ECCV 2008*, pages 495–508. 2008.

[RGT97] Y. Rubner, L.J. Guibas, and C. Tomasi. The earth mover's distance, multi-dimensional scaling, and color-based image retrieval. In *Proceedings of the ARPA Image Understanding Workshop*, pages 661–668, 1997.

[Roc71] R. Rockafellar. Integrals which are convex functionals. ii. *Pacific Journal of Mathematics*, 39(2):439–469, 1971.

[Roc15] Ralph Tyrell Rockafellar. *Convex analysis.* Princeton university press, 2015.

[Vil03] C. Villani. *Topics in optimal transportation.* Number 58. American Mathematical Soc., 2003.

[Vil09] C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften.* Springer, 2009.